Assessing the Impact of Variable Infectivity on the Transmission of HIV Among Intravenous Drug Users

Fraser Iain Lewis

Department of Statistics and Modelling Science

University of Strathclyde

Degree of Doctor of Philosophy, 2000

Copyright

Abstract

The spread of HIV and AIDS is a serious and increasing global problem with the sharing of contaminated injection equipment a primary cause of HIV infection in the developed world. Mathematical models of disease transmission allow us to assess the impact of different epidemiological and behavioural assumptions on the long term behaviour of disease.

Initially a simple deterministic model is examined which allows intravenous drug users to progress through three different infectious stages after initial infection with HIV and prior to the development of AIDS. This model is then developed to also allow contaminated injection equipment to exist in three different states of infectivity. The resulting model contains a number of parameters, which while potentially important, are extremely difficult to estimate. In response to this, several special cases are examined which represent intuitive upper and lower bounds for the spread of disease. In each case an equilibrium and stability analysis is presented. Later these special cases, together with a generalisation of them, are compared with a well established single stage infectivity model to ascertain whether the inclusion of variable infectivity increases the predicted spread of disease. We find that the impact of variable infectivity depends on a number of factors and can lead to either an increase or decrease in the prevalence of disease.

Testing drug users for the presence of HIV has been proposed as a method of reducing the incidence of HIV. Using the previously discussed upper and lower bound variable infectivity models, we examine the effect of testing addicts for HIV using a number of different infectivity assumptions. We find that under certain conditions HIV testing can be an effective control strategy against the future spread of HIV. This is followed by a short discussion of sensitivity analysis of these models. While predominantly discussing deterministic models we conclude with a brief discussion of stochastic models and demonstrate the behaviour of these models using simulation.

Acknowledgments

I would very much like to thank Dr. David Greenhalgh for his tireless help and assistance with the work contained in this thesis. I am also grateful for the financial support of the Engineering and Physical Sciences Research Council who provided a studentship to support this research.

Contents

Chapter 1

Introduction & Literature Survey

1.1 Introduction

This thesis is concerned with modelling the spread of the Human Immunodeficiency Virus (HIV) and the subsequent Acquired Immune Deficiency Syndrome (AIDS) among populations of intravenous drug users via shared injection equipment. Medical evidence suggests that once an individual has been infected with HIV the level of virus in their blood varies considerably over their infectious lifetime. Many existing models of the spread of HIV via shared injection equipment assume that the infectivity of an addict is constant throughout his or her infectious lifetime. This thesis explores the effect of a three stage infectious period with different levels of HIV infectivity in each infectious stage on long term prevalence levels of HIV and AIDS and whether control strategies such as needle exchange programs need to be reassessed. In addition we examine the testing of addicts for HIV as a potential control strategy. It has been suggested that the effectiveness of such a strategy is closely linked to the varying levels of infectivity which occur in an addict during the long AIDS incubation period.

In this first chapter we briefly outline the origins of epidemiological modelling and the effects of some of the world's worst epidemics. This leads us on to discussing the history and nature of HIV and AIDS. We then give a short summary of the state of the AIDS pandemic at the start of 1999 and in particular the role played by intravenous drug use. We then discuss some of the main modes of transmission of HIV and why populations of intravenous drug users are particularly vulnerable to HIV infection. Next we outline the main epidemiological characteristics of HIV infection before moving on to a short review of the literature concerned with the sexual spread of HIV and AIDS.

We then review a number of published articles specifically concerned with modelling the spread of HIV and AIDS among populations of intravenous drug users. Finally we conclude with an overview of the material in Chapters 2-11.

1.2 Epidemics, Plagues and Other Scourges

Of the riders of the Apocalypse, the Fourth Horseman has been the busiest. Since the beginning of history malaria has killed half of the men, women and children that have died on the planet (Nikiforuk, 1992). It has outperformed all wars, all famines and all other epidemics. The tiny proportion of Native Americans alive today is testament to the huge devastation caused by smallpox. Syphilis introduced menace to sex and people to the wig. With the plethora of sophisticated drugs available today it is easy to forget our plague-ridden past, however the Fourth Horseman rides into our lives at his convenience, with HIV and AIDS offering yet another poignant example that pestilence never rests.

Nikiforuk (1992) puts forward the case that epidemics of infectious diseases are brought about by changes in the lifestyles and habits of civilizations. For example the development of agriculture created a common market of diseases by bringing together all manner of viruses, fungi and bacteria in gardens, houses and villages. When the dog became man's best friend so too did measles; with the cow came tuberculosis and diphtheria. Rhino-viruses (the common cold) probably came riding in on a horse. Anthrax popped out of the soil. These biological collisions were probably a shock for all species involved.

Throughout history many attempts have been made to contain the spread of infectious disease, however invariably these attempts had little effect and some control strategies while well intentioned only served to exacerbate an already poor situation. The following extract is from Nikiforuk (1992) and is a good illustration of when a seemingly sensible idea can go badly wrong.

> Once DDT had established a reputation as the 'atomic bomb' of the insect world, public health officials adopted the weapon with unquestioning fervor. By the 1960's when the World Health Organization's anti-malarial campaign peaked, 76,000 tons of DDT fell on 76 countries. Although the chemical initially killed anopheles with clinical efficiency, it soon bred a stronger and more resistant adversary. At least 57 mosquitoes can now swim in DDT and

other insecticides without suffering any ill effects. Gallons of DDT sprayed randomly also produces a myriad of unforeseen health problems.

A typical case of good intentions gone awry occurred in Sarawak, part of Borneo. Here the spraying of homes with DDT not only killed the mosquitoes but cockroaches. Cats returned to the sprayed homes, ate the poisoned cockroaches and died. Free of predators, the Malaysian field rat, a carrier of plague and typhus, overran the mosquito-free villages. Fearing an outbreak of plague, the WHO eventually asked the Royal Air Force to drop cats by parachutes over the isolated villages. Fortunately for Sarawak's peasants, 'Operation Cat Drop' helped avert an epidemic of plague that DDT and malarial control had invited into their villages.

Perhaps the most clear message we can take from the past history of epidemics is that while the names and types of disease may change it seems inevitable that epidemics will continue. This means that new challenges will continually present themselves to both medical and biological researchers and the epidemiological modelling communities. The most important recent challenge is of course the HIV/AIDS epidemic and it is this challenge that the subsequent chapters of this thesis are concerned with. Before we discuss the background and particular social and biological details of the HIV/AIDS epidemic we now give a short outline of the origins of epidemiological modelling and several of the early pioneering papers in this field.

1.3 Origins of Epidemiological Modelling

We now take a very brief tour of some of the founding articles and papers of epidemiological modelling. A concise introduction to the history of epidemiology can be found in Schwager et al. (1989). The first epidemiological model was presented by Daniel Bernoulli in 1760 to the Royal Academy of Sciences in Paris. However the mathematical theory of epidemiology made no significant advances until the work of Russian physician P. D. En'ko published in 1889. En'ko constructed the first chain binomial model (wrongly attributed to Reed, see Dietz, 1988 for further details). The key concepts in the development of the mathematical foundations of theoretical epidemiology derived from the work of Major Ronald Ross who spent half of his life trying to solve "The Great Problem": how malaria infected humans (Nikiforuk, 1992). In 1902 Ross received the Nobel Prize when he identified the life cycle of the malaria parasite in birds,

and this work was described in Ross (1908), (1911) and (1915). Ross introduced the assumption that the rate of new infections is proportional both to the number of susceptibles and to the number of infectious individuals (the so-called "mass-action law"), developed the first mathematical model for the spread of vector-transmitted disease (malaria), and concluded that to eradicate malaria it was sufficient to bring the vector population below a threshold level. This was a revolutionary result in that it dispelled the belief that to control malaria one had to eradicate the mosquito population (which was an impossible task). McKendrick extended this result and in 1927 published the famous threshold theorem with Kermack, in Kermack and McKendrick (1927), which established that at least a threshold number of susceptible individuals must be present in order for an epidemic to occur.

Ross was aware of the necessity of taking into account the effects of non-homogeneous mixing, demography, geographical distribution and other factors in order to increase the predictive and explanatory power and applicability of epidemiological models. This level of detail could only be introduced by the stratification of a population into subpopulations according to specified criteria and a detailed description of the mixing between the various subpopulations. Elaboration of these ideas was greatly developed during the 1960's due to dramatic increases in venereal diseases and further expanded in response to the AIDS epidemic (which we discuss in the following section). The first mathematical model for the transmission of a venereal disease was developed by Cooke and Yorke (1973). A model for gonorrhea with an arbitrary number of randomly (proportionate mixing) interacting groups was formulated and analysed by Lajmanovich and Yorke (1976), and Hethcote and Yorke (1984) introduced concepts such as saturation and pre-emption, as well as the concept of a core subpopulation into gonorrhea analysis. This idea of a core subpopulation has been extremely important in theoretical epidemiology. Once HIV and AIDS was discovered the pace of research into modelling the spread of infectious diseases was very rapid indeed and today represents a very large body of work. Having given some background to the importance of infectious diseases and briefly outlined some of the founding work in the field of epidemiological modelling we now move on to discussing the diseases upon which this thesis focuses, namely HIV and AIDS.

1.4 A Brief History of HIV and AIDS

Between October 1980 and May 1981, five young homosexual men were treated for Pneumocystis carinii pneumonia in hospitals in Los Angeles, USA. These cases attracted attention because Pneumocystis carinii pneumonia was known to be a disease associated with immunodepression. Around the same time Kaposi's sarcoma, a skin tumour, was being diagnosed with increasing frequency in young men in New York City and in California. These observations heralded the early stages of the emergence in the USA of an apparently new disease, subsequently termed Acquired Immune Deficiency Syndrome (AIDS), (Anderson and May, 1991). The word "acquired" was used to describe the disease because unlike other immune deficiency illnesses it appeared that AIDS was an illness that you acquired from someone else as opposed to being something that happened to you, for example the taking of immuno-suppressant drugs after an organ transplant. By the autumn of 1981 the United States Public Health Service had begun efforts to try to define and understand this new disease. The hunt for a transmissible agent rapidly narrowed down to a search for a virus. In 1984 an International Committee for the Taxonomy of Viruses named the Human Immunodeficiency Virus (HIV) the aetiological agent of AIDS.

At the end of 1982 AIDS had been reported in fourteen nations worldwide. By the end of 1983 this had increased to 33 countries and two years later AIDS had been reported in 51 countries. By November 1987, 62,811 cases of AIDS had been officially reported to the World Health Organisation (WHO) from 127 countries worldwide. By the end of 1990 over 307,000 AIDS cases had been officially reported to the WHO with the actual number estimated to lie closer to one million, (MAP, 1998).

In response to the rapidly growing AIDS pandemic the Joint United Nations (UN) Programme on AIDS (UNAIDS) became operational in January 1996. UNAIDS was designed to combine the AIDS work previously undertaken by the WHO Global Program on AIDS, the UN Children's Fund, the UN Population Fund, the UN Educational Scientific and Cultural Organisation, the UN Development Program and the World Bank. UNAIDS describes itself as "the main advocate for global action on HIV/AIDS", "UNAIDS leads, strengthens and supports an expanded response aimed at preventing the transmission of HIV, providing care and support, reducing the vulnerability of individuals and communities to HIV/AIDS, and alleviating the impact of the epidemic", (UNAIDS statement to UN General Assembly on Drugs, 1998).

Table 1.1: Figures from AIDS epidemic update:1998

Global Summary of the HIV/AIDS Epidemic December 1998	
People newly infected with HIV in 1998	**5.8 million**
No. of people living with HIV/AIDS	**33.4 million**
AIDS deaths in 1998	**2.5 million**
Total no. of AIDS deaths since 1980	**13.9 million**

Table 1.1 contains figures extracted from the AIDS epidemic update: December 1998 published on the Internet by UNAIDS (www.unaids.org). The figures in Table 1.1 describe unmistakably the seriousness of the AIDS pandemic. From the beginning of the AIDS epidemic a total of 47 million people have been infected. Moreover there is evidence that AIDS is becoming more widespread, for example the number of people living with HIV/AIDS (33.4 million) at the end of 1998 has grown by 10% in just a single year. The AIDS epidemic update: December 1998 (produced by UNAIDS) states that "The [HIV/AIDS] epidemic has not been overcome anywhere. Virtually every country in the world has seen new infections in 1998 and the epidemic is frankly out of control in many places". In a speech made at The Hague on the 8th of February 1999 the UNAIDS Executive Director Peter Piot claimed that "AIDS is the single greatest threat to continued global development".

The seriousness of the AIDS pandemic is not only in the massive loss of life which ensues from it but also the severe social and economic consequences. For example UNAIDS estimates that in sub-Saharan Africa the life expectancy by 2010-2015 will be approximately 47 years of age rather than the pre-AIDS estimate of 60 years of age. In particular young people are now disproportionally affected by HIV/AIDS with around half of all new HIV infections occurring in people aged 15-24. This has very serious consequences for the prospects of future generations. The economic disruption from HIV/AIDS in developing areas such as Africa is particularly acute with the economies of whole countries becoming decimated. The World Bank Internet site, at www.worldbank.org contains a number of reports and papers on the serious economic aspects of the AIDS pandemic.

There is little doubt about the seriousness of the HIV/AIDS global pandemic, how-

ever HIV/AIDS is spread very unequally around the world both in the context of the total number of people infected and the different population groups which are worst affected. We now examine more closely the differences between the spread of HIV/AIDS in various parts of the world and in particular the significance of populations of intravenous drug users.

1.5 The AIDS Pandemic and the Role of Intravenous Drug Use

We now outline the spread of HIV/AIDS in various parts of the world and summarise the role of intravenous drug use. We first examine the epidemics in Africa, Asia and Latin America before moving on to North America, Western Europe and finally Eastern Europe. Statistics concerning the spread of HIV/AIDS throughout the world can be obtained from the UNAIDS Internet site at www.unaids.com. This site contains many reports and bulletins on the spread of HIV/AIDS, three which were of particular use were: AIDS epidemic update: December 1998; Drug use and HIV/AIDS - UNAIDS statement presented at the United Nations General Assembly Special Session on Drugs on the 9th of June 1998 and Provisional Report from MAP (Monitoring the AIDS Pandemic) entitled "The Status and Trends of the HIV/AIDS Epidemics in the World", dated June 26, 1998. All these documents are available in electronic form on the Internet from www.unaids.org. It should be noted that these sources are UNAIDS publications and as such have not been subject to an independent refereeing procedure. However we use these publications as they provide a convenient snapshot of the current global state of the HIV/AIDS epidemic and contain only observed statistics rather than projections or model based predictions.

Africa

While no country in Africa has escaped the Human Immunodeficiency Virus some are far more severely affected than others. The bulk of new infections are concentrated in East and especially Southern Africa. Sub-Saharan Africa is home to 70% of all people who became infected with HIV in 1998. During 1998 it was estimated that AIDS was responsible for 2 million African deaths. However despite the scale of death there are more Africans living with HIV than ever before: 21.5 million adults and a further 1 million children.

The vast majority of HIV infections in Africa are due to heterosexual contacts and mother-to-infant transmission. However MAP does note that the rising availability of injectable substances such as heroin, especially at new transit points for drug trafficking, creates an additional risk for HIV spread in sub-Saharan Africa. MAP summarise the spread of HIV in Africa as unabated, diverse and complex.

Asia

This region has over 60% of the world population and contains a wide diversity of HIV-related risk environments in terms of behavioural, political and cultural factors. While HIV/AIDS infection rates remain low relative to some parts of the world, well over 7 million Asians are already infected, and HIV is clearly beginning to spread in earnest through the vast populations of India and China. India alone has a population of close to 1 billion, roughly half of them in the most sexually active age group of 15-49. An estimated adult prevalence rate of about 0.6 to 1.0 percent translates to between three and five million infected persons, a figure higher than any other single country. However the distribution of HIV/AIDS in India is not uniform. For example 21 of the 31 states taken together report only four percent of the total AIDS cases. The major impact of the epidemic is being felt in Maharashtra, Tamil Nadu, Pondichery and Manipur. While the epidemic is predominantly heterosexual in nature over most of India the northeastern states have a severe epidemic among intravenous drug users. In Manipur it is estimated that the infection rate among intravenous drug users is now more than 70%.

Latin America and the Caribbean

HIV epidemics in Latin America and the Caribbean reflect the heterogeneity of HIV epidemics worldwide: they differ from country to country and within countries. The aggregate population of the forty-four countries in this region totals 476 million, 8.4 percent of the global population. An estimated 1.6 million people are living with HIV/AIDS in Latin America and the Caribbean, equivalent to 5.4 percent of the total number of people around the world living with HIV/AIDS as of January 1998 (MAP). There are a number of transmission routes responsible for the HIV/AIDS cases in this region. The dominant transmission route so far has been thorough homosexual contact and intravenous drug use, however the rising rate in women suggests that heterosexual

contact is becoming more prominent. In general the epidemics are either in an early phase or show slow growth, with several areas such as Honduras and Mexico having more established epidemics.

North America and Western Europe

In North America and Western Europe new combinations of anti-HIV drugs continue to reduce AIDS deaths significantly. However during 1998 no progress was recorded in reducing the number of new infections. Over the last decade the rate of new infections has remained stable rather than decreasing. During 1998 alone nearly 75,000 people became infected with HIV bringing the total number of North Americans and Western Europeans living with HIV to almost 1.4 million (890,000 in North America and 500,000 in Western Europe). UNAIDS states that while the epidemic is no longer out of control in these countries it has clearly not been stopped.

In Western Europe the epidemic originally occurred predominantly among male homosexuals, however subsequently the number of cases of HIV/AIDS in intravenous drug users has become dominant. UNAIDS suggests that this is due to the rapid and intensive spread of HIV through injecting drug use in populations of south-western Europe, particularly Spain, Italy and more recently Portugal. In North America and Western Europe heterosexual contact accounts for an increasing proportion of HIV/AIDS cases though this is still small in comparison to the number of cases in either homosexual males or intravenous drug users.

Eastern Europe

Until 1995 Eastern European countries including the Asian republics of the former Soviet Union reported few HIV cases, these were mostly among homosexual men. Since 1995 HIV has spread very rapidly among intravenous drug users in cities of several countries including the Ukraine, Belarus, Moldova and the Russian Federation. Epidemics among drug users are also emerging in the Caucasus, the Baltic States and in Kazakstan in Central Asia. UNAIDS and WHO estimates that the total number of infections may have risen from less than 30,000 in 1995 to more than 270,000 in 1998. According to MAP the main factors fueling the HIV epidemic among intravenous drug users are increased drug demand and supply, migration and widespread local drug production. It is estimated that Russia alone has over 1 million intravenous drug users.

Summary of the Importance of Intravenous Drug Use

To summarise we have that the truly massive number of HIV/AIDS cases in areas such as sub-Saharan Africa and India arise from heterosexual contact. However in many parts of the world injecting drug use is the major mode of HIV transmission. Moreover UNAIDS claims that "injecting drug use plays a critical role in how and when the HIV epidemic starts in a particular region and how it continues to unfold". More than half of all reported cases of HIV/AIDS in the Russian Federation have been in intravenous drug users. The UNAIDS statement to the United Nations General Assembly Special Session on Drugs (9th June 1998) claims that in the world today there are at least 5.5 million and possibly up to 10 million intravenous drug users, ranging across 128 countries and territories, up from around 80 six years ago. In the USA alone it is estimated that 700,000 people inject illegal drugs intravenously, whilst Russia has seen a massive 20-fold increase in the number of intravenous drug users since 1990 taking the estimated total to 1 million users. In short the study of the spread of HIV and AIDS among populations of intravenous drug users can easily be justified as relevant and potentially important in helping combat the global spread of HIV and AIDS.

1.6 The Transmission Aspects of HIV

Before we discuss why drug users are a group at particularly high risk of infection from HIV, we first outline the various ways in which HIV can be transmitted other than through intravenous drug use. Other than drug related HIV infection there are three main modes of HIV transmission, these are infection from receiving a contaminated blood transfusion (or other related blood products such as plasma), infection through sexual contact and infection from mother to infant (perinatal transmission). Some other speculative forms of HIV transmission have appeared in the press, such as infection from insects and close (non-sexual) personal contact. There is no evidence to suggest that insects carry HIV or that close non-sexual personal contact carries any risk of HIV transmission, (Friedland and Klein, 1987). We now briefly outline each of the three main (non-drug related) modes of HIV infection.

Firstly we examine the transmission of HIV via the transfusion of blood and blood products. Persons who have acquired AIDS through the transfusion of infected blood or blood products represent a small but important proportion of the total number of AIDS cases. Up to January 1987, 2 percent of adults and 12 percent of children with

AIDS in the USA were believed to have acquired the disease in this manner (CDC, 1987). Transfusion related AIDS has provided a number of important insights into HIV transmission. Arguably the most important of these is the information available concerning the duration of the AIDS incubation period. This mode of transmission represents a single point in time of inoculation as opposed to probable repetitive exposures over time characteristic of other modes of infection. HIV infection by transfusion can therefore provide accurate information as to the likely length of the AIDS incubation period. In response to the risk of transmission of HIV by blood and blood products, the screening of donated blood and plasma for antibody to HIV began in 1985. This reduced the risk of infection from blood transfusion and blood products to a very low level, from between 1 in 100,000 to 1 in 1,000,000, (Friedland and Klein, 1987). Friedland and Klein argue that other methods of HIV transmission are not susceptible to such simple technical solutions as blood screening.

HIV is fundamentally a sexually transmitted virus, which is transmitted by both homosexual and heterosexual activity. The risk of acquiring HIV infection from a single or several sexual encounters with an infected person is not known. However it is known that other sexually transmitted diseases such as syphilis and gonorrhea carry a substantial risk from a single encounter, which increases with repeated encounters. Available information suggests that HIV is less easily transmitted and in studies of steady heterosexual partners the overall rates of infection ranged from 7 to 68 percent, Fischl et al. (1987), Redfield et al. (1985), Kreiss et al. (1985) and Peterman et al. (1986). Although the risk of transmission is substantial in all populations, most regular sexual partners remain uninfected, (Friedland and Klein, 1987). However it is interesting to note that studies of heterosexual and homosexual communities (BMBR, 1987) reveal an important difference between these two risk groups. There appears to be an eight to twenty-fold difference in the average relative rate of acquisition of partners between heterosexuals and homosexuals, (Anderson and May, 1991). Irrespective of the relative magnitudes of other factors that determine the rate of transmission of HIV, this factor by itself suggests that the spread of HIV in heterosexual communities in developed countries will be much slower than that observed in homosexual communities.

The third main mode of HIV transmission is from a mother to her offspring. In the USA, 80 percent of children with AIDS are known to have a parent who has AIDS or is at risk of the disease and presumably infected with HIV (CDC, 1987). Since pediatric AIDS is closely linked to maternal infection, it is not surprising that its demographic

features in the USA parallel those of AIDS in women and are closely tied to intravenous drug use. HIV may be transmitted from infected women to their offspring by three possible routes: to the fetus in utero through the maternal circulation, to the infant during labour and delivery by inoculation or ingestion of blood and other infected fluids, and to the infant shortly after birth through infected breast milk, (Friedland and Klein, 1987). Presumably in an infant infection may occur by any combination of these routes, but the relative efficiency and frequency of each route is not known. The final and arguably the most efficient route of HIV transmission is in an intravenous drug using environment where infectious needles are shared among addicts. It is this mode of transmission with which the models in this thesis are concerned and it is this we now discuss.

1.7 Why Drug Users are at High Risk from HIV/AIDS

Having illustrated that the spread of HIV/AIDS among intravenous drug using populations is a widespread problem we now briefly discuss why this particular group are at such a high risk of infection from HIV/AIDS. There are a number of ways in which HIV can be transmitted from addict to addict in an intravenous drug using environment. The most obvious method of transmission is when an addict infected with HIV uses a needle and leaves a residue of HIV infected blood in the needle. The next user (or sharer) of the needle is then exposed to this infectious blood and may become infected. UNAIDS claims that of all the different ways that the HIV virus can be passed on, directly injecting a substance contaminated with HIV into the blood stream is by far the most efficient, much more so in fact than through sexual contact. This raises the obvious question of why addicts share needles when the risks attached are so severe. However it is not merely the direct sharing of needles which transmits HIV among addicts, there are a number of related methods of infection concerned with preparing the drug prior to injection which are often not perceived as risky by addicts but which do carry the possibility of HIV infection, (Hunter et al., 1995). We now briefly discuss why addicts share needles and what additional risks are associated with drug injection practices.

1.7.1 Needle Sharing

It is not clear why addicts share needles (Hay, 1999), however there are a number of reasonable explanations, perhaps the most obvious of which is the lack of access to sterilised needles. In many parts of the world, in particular much of the USA, it is illegal to purchase or possess drug injecting equipment without a valid prescription, and this is backed up by stiff criminal penalties. The motivation behind this policy is to use legislation to decrease the availability of injection equipment, and thus prevent people being able to inject drugs and therefore cause the size of the IVDU (intravenous drug using) population to decrease. However it could be argued that there is a serious incompatibility between policies to control drug use and policies to control the spread of HIV among intravenous drug users.

Intuitively it is not hard to see that restricting the supply of uncontaminated injection equipment in this way could produce a vicious cycle in which the prevalence of AIDS among the general population increases (and will continue to increase) as a direct result of such a policy. Intravenous drug use is socially undesirable and not just because of the AIDS issue. Drug use is often directly linked to crime, poverty and general social deprivation. This provides motivation for governments to pass laws which make possessing injecting equipment a serious crime in an effort to restrict the population size of IVDUs. However the intravenous drugs used by addicts, such as heroin, are highly addictive and therefore rather than reduce the size of the population all this may achieve is to force addicts to share needles with greater frequency. With fewer needles available, a needle is shared between more people and hence is more likely to become contaminated with HIV as time progresses. Therefore the prevalence of HIV/AIDS in the population of IVDUs increases and since the IVDU population interacts sexually with the general population this causes an increased threat of HIV infection to society at large. This in turn puts additional pressure on governments to decrease the IVDU population by further restricting access to injecting equipment and so on.

The type of needle sharing environment which has emerged as a result of making the possession of injection equipment illegal is particularly hazardous. Where the possession of injection equipment is illegal it is common for drug addicts to frequent "shooting galleries" to share needles. A "shooting gallery" is a place where drugs are purchased and where drug injection equipment is sequentially rented to users (Kaplan, 1989a). This is a very hazardous environment as large groups of addicts inject with the

same needles which are often used without effective cleaning and where the injection equipment is not regularly replaced (Des Jarlais et al., 1987). Hay (1999) describes various types of shooting galleries, such as the residential gallery where long and short term residents share injection equipment with non residents. A particular danger of visiting a shooting gallery is the lack of knowledge that addicts have about the behaviour of each other, and it is quite possible that a number of the addicts who frequent the gallery may be infected with HIV. It is commonly accepted that a shooting gallery is one of the most dangerous environments for the spread of HIV among a population of intravenous drug users. For example Kaplan and O'Keefe (1993) found that in a sample of needles taken from a shooting gallery in New Haven, Connecticut, USA, 44 out of 48 needles tested positive for HIV. Hence it seems that a policy which restricts the supply of sterilised needles can have a very serious effect on the spread of HIV among an intravenous drug using population (by forcing the widespread sharing of needles).

A lack of availability is not the only reason why addicts share needles. Needle sharing still occurs in areas where needles and injection equipment are freely available (Kretzschmar and Wiessing, 1998). A common type of needle sharing is called "friendship networks", these are where addicts share needles with close friends or sexual partners. The addicts themselves do not class this as sharing and as such see no risk of HIV infection from this behaviour (McKeganey and Barnard, 1992). However as with any needle sharing this behaviour does carry a real risk of HIV infection.

A further reason why addicts share needles may be because they do not recognise the risks involved. This argument is perhaps less convincing in the late 1990's than in the early 1980's when HIV was first diagnosed, since there has been an increased public awareness to the dangers of HIV and AIDS in recent years. However that is not to say that all new drug users (particularly young ones) recognise the risks of needle sharing.

1.7.2 Drug Preparation and HIV Infection

We now examine the risks of HIV infection associated with the manufacture and preparation of drugs prior to injection. Before heroin can be injected it must first be dissolved in water. This can be done in small vessels such as bottle caps known as "cookers" which are heated, (Friedland and Klein, 1987). Some users will place cotton in this vessel to filter the dissolved drug. This filter may be used several times, hence if an infectious needle has been used with this filter then the cotton may be infectious and

a possible source of transmission of HIV. In addition when drug users want to share drugs which have been jointly purchased, instead of dividing up the raw drug it may be easier to divide up the dissolved solution. One method of dividing up this solution is known as "front-loading", (Grund et al., 1990). This involves drawing all the solution into a single syringe then injecting some of the contents into other syringes. If the initial syringe is infectious then there is a risk of contaminating subsequent syringes.

In Eastern European countries and the Russian Republics the probability of transmission of HIV through intravenous drug use is particularly acute. In these areas most intravenous drug users use home-made opiates, and HIV risk appears to depend on specific drug preparation and distribution patterns (MAP, 1998). Needle sharing has been reported from all categories of injecting drug users but MAP believes that the use of home-made opiates is particularly risky. The reason for this is that home-made opiates may be prepared with contaminated equipment and ready-made drugs are sold in used and potentially contaminated syringes or other containers. In addition, according to MAP, human blood is often added to the drug solution as a cleansing agent during the production process (often coming from an HIV infected individual). These factors together create a particularly hazardous environment for intravenous drug users in Eastern Europe and the Russian Republics.

We have shown the importance of drug using populations in the spread of HIV/AIDS and illustrated why these populations are so vulnerable. We now take a brief look at some of the epidemiological characteristics of HIV infection which are relevant to modelling the spread of HIV and assessing the public health issues involved in the AIDS pandemic.

1.8 Epidemiological Characteristics of HIV Infection

We first give some biological background on HIV before discussing how it manifests itself in humans. HIV is a retrovirus with morphological, molecular, and biological characteristics that have led to its proposed classification with the pathogenic animal lentiviruses (Anderson and May, 1991). Recent work has demonstrated that other primate species, notably the African Green monkey, may be symptomless carriers of a related retrovirus that can also infect humans, apparently without causing disease (Kanki et al., 1985). Since the discovery of HIV in 1983, the progress of research at the molecular and cellular levels of study has been very rapid indeed (Anderson and May,

1991). However in sharp contrast to the current understanding of the genetic structure and function of the virus, and the pathology induced by infection, comparatively little is known concerning the epidemiological and transmission dynamics of HIV.

Once an individual is infected with HIV evidence suggests that the virus persists for the life of the individual (Weiss, 1985). After the initial infection with HIV the incubation period until the development of AIDS can be a few to many years with median estimates at approximately 9.8 years (Longini et al., 1989). The duration of the incubation period from initially being infected with HIV until the development of AIDS has been subject to major changes in recent years due to combination therapy drugs. For example UNAIDS reports that the death rate due to AIDS in the USA is the lowest in a decade (UNAIDS AIDS epidemic update, 1998), this is due to the ability of new combination drug therapies to substantially delay the onset of full blown AIDS, perhaps indefinitely. This situation poses a new problem for public health policies since the total number of people living with HIV has greatly increased in certain parts of the world due to the lowering of the AIDS mortality rate. However it should be noted that in terms of the global HIV/AIDS situation the proportion of infected individuals who have access to these new combination drug therapies is tiny ($< 5\%$) and they are very costly.

Returning to the incubation period of AIDS, we find that after initial infection the individual enters an incubation period prior to primary HIV infection. This may last a few weeks before progressing to the primary HIV stage which may last a few days to weeks. Next the infected individual enters a stage with no symptoms (the asymptomatic stage) which may last a few to many years before entering the AIDS Related Complex (ARC) stage. This is where the infected individual is classed as having a syndrome of chronic unexplained lymphadenopathy and persistent depletion of T-helper cells. This stage may last a few to many years, it is sometimes called the pre-AIDS stage of infection. Finally the infected individual enters the stage of full blown AIDS in which life expectancy is approximately 0.5-1.0 years (Anderson and May, 1991).

In studying the transmission dynamics of HIV infection it is important to ascertain the relationship between the incubation period of the disease, and the duration and intensity of infectiousness over this incubation period, (Anderson and May, 1991). Anderson and May suggest that whilst data in this area is very sparse, a tentative hypothesis from the studies which have been conducted suggests that there are two phases of peak infectivity during the long incubation period of AIDS, one lasting for

a few months to a year or more following initial infection, and the other just prior to the onset of AIDS. In the intervening periods, infectiousness of the patient may well be very low. This is a very similar situation to that found by Peterson et al. (1990) and Seitz and Müller (1994).

1.9 Review of Modelling HIV/AIDS

Having briefly examined several epidemiological aspects of HIV/AIDS we now turn our attention to published literature concerned with modelling the spread of the disease. There are a vast number of articles relating to modelling the sexual spread of HIV/AIDS. In this section we do not seek to provide a thorough review of all this literature, but rather provide a flavour of the kind of modelling work which has been undertaken. Several extensive reviews of the literature concerned with modelling HIV/AIDS can be found in Abrams (1987) and Schwager et al. (1989). The amount of published work concerned with modelling the spread of HIV in populations of intravenous drug users via needle sharing is much more sparse, and we review some of this work in detail in the following section.

May and Anderson (1987) discuss the transmission dynamics of HIV. This paper first reviews data on HIV infections and AIDS disease among homosexual men, heterosexuals, intravenous drug users and children born to HIV infected mothers. All the information currently available (prior to 1987) concerning the distribution of incubation times that elapse between HIV infection and the appearance of full blown AIDS, the fraction of those infected with HIV who eventually develop AIDS, the time-dependence patterns of infectiousness and the distribution in the rates of acquiring new sexual or needle sharing partners is used by May and Anderson to develop models of the transmission dynamics of HIV. They begin with deliberately oversimplified models and progress (on the basis of understanding thus gained) to more complex models. Later in this paper the degree to which sexual or other habits must change in order to bring the basic reproductive number of HIV infections to below one is discussed. May and Anderson conclude that the epidemiology of HIV has many unusual features, and making accurate long term predictions about the prevalence of HIV and AIDS within any particular group requires an understanding of the nonlinear dynamics of HIV transmission and extensive reliable data. Given the current uncertainties about so many of the biological and sociological aspects of HIV transmission, May and Anderson

believe that it is sensible to begin by exploring relatively simple models that caricature the transmission dynamics, with a view to understanding the qualitative features of the HIV/AIDS epidemic.

Blythe and Anderson (1988a) explore the effect of the incubation period and infectious periods in models of the transmission dynamics of HIV. A series of risk (or hazard) functions are used to describe variation in the incubation and infectious periods of HIV. Four forms of distribution are considered, namely, exponential, Weibull, Gamma and rectangular. Models of the transmission dynamics of the virus encompassing different assumptions concerning the distributed incubation and infectious periods are analysed, and their properties compared by steady state and local stability analyses and numerical methods. Blythe and Anderson find that which distribution should be used for the incubation period of the disease depends mainly on three factors. Firstly the distribution chosen should provide a general empirical description of the available epidemiological data. Secondly the risk function should again be chosen to mirror additional biological, immunological, or clinical data concerning the factors that determine progression to the disease state of AIDS. Finally, the ease with which analytical and numerical results can be obtained from the model containing the distribution should be borne in mind. Each of the four distributions examined in this paper yielded very similar results with respect to the steady state behaviour of the respective models and their local stability properties. Moreover Blythe and Anderson find that the full nonlinear behaviour of the different models was also not greatly different for the incubation period assumptions. Hence they conclude that provided that the mean of the distributed incubation period is fixed at the observed value then almost any distribution of approximately the right shape will suffice.

Anderson (1988) examines two topics of particular relevance to the study of the transmission dynamics of HIV, namely variability in incubation and infectious periods and heterogeneity in sexual activity within homosexual and heterosexual communities in the UK. Anderson et al. (1986) examined the important related issue of the relationship between the infectious and incubation periods of AIDS. Anderson (1988) develops simple deterministic models to describe two episodes of infectiousness during the long and variable incubation period of AIDS. These simple models are used to address two specific problems, firstly the extent to which the variability in infectiousness affects the early stages of the epidemic, and secondly the measurement and analysis of heterogeneity in sexual activity (defined as the rate of sexual partner change per unit time) within

homosexual and heterosexual communities. The main conclusion of Anderson (1988) is that much more research is needed, particularly in terms of collecting good quality data before any firm conclusions can be drawn.

Blythe and Anderson (1988b) model the sexual transmission of HIV using a proportionate mixing single sex model in which sexual activity (new partners per unit time) is defined as a continuous variable in a set of integro-partial-differential equations. A discrete activity class approximation model is developed by matching equilibrium state and rate variables as closely as possible with the continuous variable model, and consists of only ordinary differential equations. Blythe and Anderson examine the relationship between the discrete and continuous variable models using both numerical and analytical studies in order to evaluate the accuracy of the approximation.

In Isham (1988) mathematical modelling of transmission of infection in the context of the AIDS epidemic is reviewed. This paper first develops a simple homogeneous deterministic epidemic model which is adapted to follow the characteristics and peculiarities of HIV/AIDS. This model contains only five parameters: the probability of transmission from infective to susceptible per partnership; the rate of partner change; the probability that an infective will develop AIDS; the mean incubation period for AIDS patients and the mean infectious period for HIV infected patients. Isham shows that this model is reasonably applicable to a homogeneous, highly active homosexual community in the short term when the community can be assumed to be closed. This model is then extended to allow for immigration and emigration, variation in the rate of partner change and a non-exponential incubation period. This extended model is then applied to a very heterogeneous heterosexual population. Isham notes that as the model becomes more complex it becomes important to identify which sources of variation are critical in their effect on the spread of infection, and those which can effectively be assumed to be constant in order to obtain a "bróad-brush" picture, which is the most that is needed for many practical purposes. Isham concludes with the remark that due to the lack of reliable data available from which to estimate model parameters, the main practical motivation for studying HIV/AIDS models is to gain understanding of the most important factors influencing the spread of HIV infection rather than predicting future HIV/AIDS trends.

Stanley and Hyman (1988) contains an extensive discussion of the use of mathematical models to understand the AIDS epidemic. This paper contains a justification of why models which are constructed by curve fitting methods (such as regression models) can-

not be used reliably for long periods of time, nor can they provide an understanding of the interactions that lead to the spread of the epidemic. During the long asymptomatic period after infection with HIV, changes in the environment of viral transmission occur continuously, causing complex interactions. Stanley and Hyman argue that only models which are founded on the transmission mechanisms of HIV can show how the early infection of high risk groups, behavioural changes, and future medical advances such as treatments and vaccines will affect the future course of the HIV/AIDS epidemic. Stanley and Hyman briefly discuss a number of important modelling factors such as age, sexual activity and drug use as well as a number of other cofactors.

Kaplan (1989b) discusses models of sexual mixing and HIV transmission and in particular whether a simple model, while technically poor, can still provide useful insight into potential control strategies for combating the spread of HIV and AIDS. An important point which Kaplan draws attention to is that of how precise a model must be to provide guidance for policy makers. Kaplan uses a highly tailored example of a model which assumes random mixing, which while not an accurate assumption still identifies useful control strategies. The main point Kaplan makes is that a model which is oversimplistic may still contain enough of the basic features of the spread of disease to point policy makers in the correct direction.

Kaplan and Lee (1990) examine HIV modelling assumptions in order to create a model which offers a worst case scenario. Specifically they examine the maximum number of infections that could occur under any feasible mixing pattern. Such worst case results are of special interest to decision makers who must prospectively evaluate the consequences of planned public interventions. Kaplan and Lee derive upper bounds for the maximum number of infected persons possible under endemic steady state conditions within the workings of a heterogeneous mixing model of HIV transmission. Two examples are presented that utilise this bound for a range of parameter values reasonably descriptive of HIV/AIDS. These examples show that the worst case number of infected persons in the endemic steady state can be well within 10% of the number of infected persons that would result from random mixing.

Thieme and Castillo-Chavez (1989) examine the effects of infection-age-dependent infectivity in the dynamics of HIV/AIDS. They discuss the epidemiological and behavioural factors crucial to the dynamics of HIV/AIDS and include long and variable periods of infectiousness, variable infectivity, and the processes of pair formation and dissolution. Models which explore the effect of a long AIDS incubation period are dis-

cussed in Castillo-Chavez et al. (1989a), (1989b) and (1989c). In Thieme and Castillo-Chavez (1989) the role of variable infectivity in combination with a variable incubation period is examined in a homogeneously mixing population. It is shown that if the basic reproductive number is less than one then the disease dies out while if it is greater than one then the prevalence of disease converges to or oscillates around a uniquely determined nonzero equilibrium. Thieme and Castillo-Chavez observe that oscillations about the nonzero equilibrium cannot be excluded in all cases and may occur if the variable infectivity is concentrated at an early part of the AIDS incubation period.

Sattenspiel et al. (1990) discuss a model for the spread of HIV among a population of male homosexuals. In this model the population is divided into five groups on the basis of degree of sexual activity. Within each group the individuals are classified as susceptible, infective or removed due to the development of full blown AIDS. The infective individuals are further subdivided into four stages of infection. Analyses of this model address two questions with regard to the spread of HIV: firstly how the sexual activity of an individual affects their risk of infection, and secondly how assumptions relating to how different population groups mix affect both the risk to individuals and transmission through the population as a whole. Results from analyses using a number of different parameter estimates show that increased levels of sexual activity increase the likelihood that an individual will become infected. In addition the initial spread of the disease is markedly changed by variation in the amount of contact among individuals from different subpopulations. The steady state incidence of the disease is not markedly changed by variation in the contact patterns, but the size of the steady-state population and therefore the number of infected individuals in the population does vary significantly with changes in the degree of mixing among subpopulations.

Sattenspiel and Castillo-Chavez (1990) explore the effect of the spread of HIV among a male homosexual population with a specific view to determining the effect of environmental context and social interactions. They split the homosexual population into different groups according to sexual activity but additionally allow individuals to temporarily take on characteristics of a different risk group (due to the current environment entered into by the individual). This paper concludes that when the goal of modelling is to increase understanding of the transmission system of HIV, then the decision to incorporate context effects must be evaluated with reference to the focus of the particular model being used. Specifically models which focus on variability in parameters that would be strongly affected by the conditions operating at the time of contact must

consider the effects of context.

Lin (1991) examines a proportionate mixing model first developed and discussed by Anderson et al. (1986) and subsequently extended by Castillo-Chavez et al. (1989a). Lin's model consists of a sexually active population divided into n subpopulations where each subpopulation is re-divided into three epidemiological classes: susceptibles, HIV infecteds and those with full blown AIDS. Lin shows that when the disease free equilibrium is unstable the model can have multiple positive endemic equilibria. This together with results from Castillo-Chavez et al. (1989b), Castillo-Chavez et al. (1989c) and Huang (1989) indicate that circumstances exist for which the stability of the disease free equilibrium cannot be chosen as a threshold condition, which is contrary to the homogeneous mixing case. Lin shows that the proportionate mixing model is very sensitive to the HIV transmission rates and considerable variation among the transmission rates could cause the model to have multiple endemic equilibria, in which case the prediction of the spread of HIV/AIDS becomes very difficult.

Huang et al. (1992) discuss stability and bifurcation for a multiple group model of the dynamics of HIV/AIDS transmission. A multi-group epidemic model is used with a variable population size. They show that even in the case of proportionate mixing, multiple endemic equilibria are possible. The basic reproductive number is identified and it is shown that this governs the stability of the disease free equilibrium. An interesting result of this paper is the importance of using a variable population size in disease dynamics. Earlier models of sexually transmitted diseases assumed that the population and the subpopulations under consideration had a constant number of individuals. Under this assumption a very specific type of mixing is introduced, a mixing that is independent of the population dynamics. Models with varying population size make it possible to study the effects of mixing in population and disease dynamics effectively. The results of this paper imply that even in the case of proportionate mixing using a variable population size can generate multiple endemic equilibria. Huang et al. suggest that multiple endemic equilibria may occur for realistic parameter estimates and as such it seems possible that if this behaviour turns out to be generic then the potential for the use of models similar to that discussed in this paper will be very limited. Related work which discusses mixing and disease dynamics is Castillo-Chavez (1989b), Anderson et al. (1989), Anderson et al. (1990) and Blythe et al. (1991).

1.10 Review of Modelling HIV/AIDS in IVDU Populations

As mentioned in the previous section there are a considerable number of articles relating to modelling spread of HIV/AIDS, however much of this work has focused on the sexual spread of the disease rather than the spread via drug use. Literature on mathematical modelling of the spread of HIV amongst injecting drug users (IVDUs) is much more sparse, (Greenhalgh and Hay, 1997).

To our knowledge the first attempt at modelling the spread of HIV via needle sharing among a population of intravenous drug users is due to Kaplan (1989a). This is a pioneering paper as the model featured in it has been the basis for much of the literature on modelling the spread of HIV among intravenous drug users. Kaplan examines the spread of HIV via shared drug injection equipment in "shooting galleries". As mentioned previously a shooting gallery is a place where addicts go to both purchase and inject drugs. It is thought that the risk of HIV infection from needles in a shooting gallery is particularly high.

Kaplan models the spread of HIV using a simple system of two coupled differential equations. One differential equation models the fraction of infectious addicts in the population, and the other the fraction of infected needles circulating in the population (or equivalently the probability that a randomly chosen addict is exposed to HIV via sharing injection equipment). This model depends on quantities such as the rate at which injection equipment is shared, the ratio of addicts to injection equipment in the population, the probability of transmission of HIV from using infectious injection equipment, the likelihood that infectious equipment is rendered virus free by an uninfected user and the duration of needle sharing activities by HIV infected addicts.

Kaplan first demonstrates that in order for the prevalence of disease to reach an endemic equilibrium in the addict population, R_0, the number of secondary infections from a single infectious addict (assuming an otherwise totally susceptible addict and needle population) needs to be greater than unity. By a secondary infection we mean addicts infected directly by needles infected directly by the given addict. Kaplan uses his model to examine the effect of the "gallery ratio", that is the ratio of addicts to needles in a shooting gallery, on the long term prevalence of HIV in the addict population. Kaplan demonstrates that the "gallery ratio" has no effect on the long term prevalence of HIV but does have an effect on the speed at which HIV spreads

among the addict population. He then extends his model to account for heterogeneity in the frequency at which addicts share needles. Evidence suggests that the distribution of addicts who visit shooting galleries is highly skew with a large number of addicts visiting very infrequently and a small minority who visit with very high frequency (Page, 1990). Kaplan incorporates this effect into his model by introducing a further equation which deals with the likelihood that any particular addict who attends a shooting gallery at time t is HIV positive (this takes into account heterogeneity in the rate at which addicts share needles or equivalently visit a shooting gallery). Kaplan notes that an effect of introducing heterogeneity into the needle sharing rate (or shooting gallery visiting rate) is to increase the basic reproductive number.

Kaplan next examines the effect of the cleansing of injection equipment. He looks at the impact of needle cleansing *after* use and finds that this can have a significant effect in lowering the long term prevalence of HIV in the population. With hindsight it is rather strange to assume that addicts would clean a needle *after* use rather than *prior* to use. Cleaning prior to use can make a possibly infectious needle virus free for the addict to use now, whereas cleaning after use serves only to protect the next user. In Caulkins and Kaplan (1991), the original model from Kaplan (1989a) is used as a basis for assessing the impact of AIDS on the size of intravenous drug using populations in the USA. This paper highlights the problem of assessing policies to reduce the number of intravenous drug users in the USA in the light of deaths from AIDS. Caulkins and Kaplan used their model to show that AIDS could reduce the size of drug injecting populations by more than 50%. This has major implications for how policies designed to reduce the number of intravenous drug users should be evaluated.

In Kaplan and O'Keefe (1993), Kaplan's original model is extended to incorporate the impact of cleansing or bleaching of injection equipment prior to use and to allow needles to be removed from the population and be replaced by unused (and obviously uncontaminated) needles. In addition this model now assumes that an infectious needle can never be rendered virus free after use by an uninfectious addict. Kaplan and O'Keefe used their model to evaluate a pilot needle exchange program set up by the New Haven Health Department in Connecticut, USA. This scheme was initially set up after a survey of addicts found that the sharing of injection equipment was largely due to the lack of availability of unused needles in New Haven. This needle exchange program was set up to allow an addict to exchange a used needle for an unused needle on a one-for-one basis. Each addict was registered anonymously with the scheme and on

the first visit to the scheme received a single unused needle. To enable this scheme to be evaluated Kaplan and Heimer (1992a, 1992b) developed a needle tracking and testing system. This involved labelling all the program needles with serial codes so that it could be ascertained how long needles remained in circulation in the population. In addition when each needle was returned to the program it was tested for the presence of HIV. By interviewing the participants of the needle exchange program (using self assessment questionnaires) and using information from the needle tracking and testing scheme Kaplan and O'Keefe could estimate the parameters in their model and thus examine the long term effect of maintaining the pilot needle exchange program. Kaplan and O'Keefe examined the decrease in long term prevalence of HIV in addicts in their model when needles circulated among the addict population for approximately 17.8 days on average (the post-needle exchange circulation time) compared to circulating indefinitely (the assumed pre-needle exchange circulation time). This represented the effect of introducing and maintaining a needle exchange program (whose primary purpose was to lower the time a needle spends in circulation). Kaplan and O'Keefe's model suggested that the long term prevalence of HIV could be decreased by up to 33% by implementing and maintaining a needle exchange program.

Kaplan and O'Keefe also conducted a survey of the intravenous drug using population to determine whether the introduction of the needle exchange program had resulted in an increase in overall drug use (in other words increased recruitment from the general population). This survey suggested that no increase in drug use had occurred which was a particularly important result as this had been a major criticism of introducing such a program. As a result of the evaluation study by Kaplan and O'Keefe the pilot needle exchange program in New Haven was extended making this scheme the first federally funded study of needle exchange in the USA and 20 months later the purchase of hypodermic syringes was legalised in Connecticut. The success of this study was (at least in part) responsible for altering legislation in New York City, California and Massachusetts to allow needle exchange programs to be developed.

In a couple of later articles, Kaplan (1994, 1995) formulates a circulation theory of needle exchange that highlights the impact such programs have on the behaviour of needles. In essence Kaplan argues that the shorter the duration of time a needle circulates in the population the less likely it is to become infectious. This is where needle exchange programs play an important role by reducing the mean duration any given needle spends in circulation. In these articles Kaplan uses a different modelling

approach from his previous work in an effort to establish empirically the decrease in the long term prevalence of HIV/AIDS caused by introducing a needle exchange program.

One of the potential flaws in the model used in Kaplan and O'Keefe (1993) was that most of the parameter estimates required by this model could only be obtained subjectively from the addicts themselves. For example this model requires an estimate of the frequency at which addicts share needles and the proportion of times addicts cleaned needles prior to use. These highly subjective behavioural parameters can only be obtained by interviewing addicts, and Kaplan (1994) believes that parameter estimates obtained in this way could be biased towards practices which are safer than those actually undertaken (although there may also be a possibility of them being biased in the other direction). To overcome this problem he devised a model of the behaviour of needles which contained only three parameters, the rate at which an uninfected needle becomes infected, the rate at which an infected needle becomes uninfected and the average duration a needle spends in circulation. Each of these parameters could be estimated objectively using data from the needle tracking and testing scheme. Kaplan then used this model to estimate the reduction in the long term prevalence of HIV in needles due to the decrease in needle circulation time caused by the introduction of a needle exchange program. He found that the reduction in the long term prevalence of HIV in needles was approximately 33% which suggests that the number of new infections (in addicts) from these needles will also be reduced by 33%. This is further good evidence of the potential benefit of needle exchange programs.

While the model used in Kaplan (1994) makes very good use of the data available, and contains arguably some of the most accurate parameter estimates in any HIV model it is not without problems. For example the model assumes that the prevalence of HIV in addicts is constant and hence examines only the behaviour of needles. This is very unlikely since the prevalence of addicts and needles are inextricably linked. However that said, the parameter estimates used by Kaplan suggests that his model reaches a steady state very rapidly (in approximately one year). Also the timescale on which the addicts inject, typically of the order of a few days, is very much smaller than that of the other epidemiological and demographic processes involved, with the possible exception of the needle exchange rate. Hence it could be argued that provided the epidemic is not growing rapidly (and thus causing the prevalence of HIV in addicts to change rapidly) then this model may be fairly reasonable.

We have briefly discussed some of the work by Kaplan and whilst there is little doubt of the important practical significance of this research there are a number of shortcomings in the models which Kaplan uses. A number of papers have been devoted to extending the original model by Kaplan (1989a) to reflect more realistically the behaviour of a population of intravenous drug users. Several substantial extensions have been investigated by Greenhalgh, and Greenhalgh and Hay and it is this work we now discuss.

One of the major deficiencies in Kaplan's original model was that it assumed a single homogeneous population of addicts and needles. As a first investigation into the spread of HIV via shared injection equipment this is entirely reasonable, however a more realistic assumption is that the drug using population contains many sub-groups, each with different attributes and behavioural characteristics. For example in a large metropolitan area such as New York City which has an estimated 200,000 drug users, (Ginzburg, 1984 and Drucker, 1986), it is highly likely to be the case that many shooting galleries exist, where each of these may have a different composition of drug users (in terms of the frequency of sharing and needle cleansing). In addition it is unrealistic to assume that each of the 200,000 addicts is equally likely to visit each shooting gallery as it seems reasonable to expect addicts to visit mainly shooting galleries in their own locality. Moving from Kaplan's original model to the highly heterogeneous situation just described is an extremely important step since this reflects much more accurately the behaviour of drug using populations throughout the world.

Greenhalgh (1996) extends Kaplan's original model to incorporate variability in the rate at which addicts visit shooting galleries (or equivalently the frequency of needle sharing), the choice of shooting gallery, and the probability that addicts clean needles prior to injection. This model consists of a system of differential equations which Greenhalgh analyses to investigate the conditions necessary for the disease to die out or remain in the population. He found that (as implied by Kaplan's original model) a necessary and sufficient condition for HIV/AIDS to become endemic in the population is that the basic reproductive number exceeds unity, moreover if $R_0 > 1$ then the disease will tend to a unique endemic equilibrium and if $R_0 \leq 1$ then the disease will die out in all addicts and all needles.

In Greenhalgh (1997) it is demonstrated that heterogeneous mixing of addicts and needles (as modelled in Greenhalgh, 1996) can give rise to higher long term prevalence levels of HIV/AIDS than homogeneous mixing. This is a potentially important result as

Kaplan (1989a) argues that homogeneous mixing is always a conservative assumption and overestimates the equilibrium prevalence of HIV amongst addicts. The results of Greenhalgh (1997) imply that it may not be suitable (as previous thought) to examine the spread of HIV among populations of drug users using simple homogeneous models.

In Greenhalgh and Hay (1997), the effect of more of the simplifications made in Kaplan's original model are investigated. Kaplan assumed that an infectious drug user always leaves a needle infectious after use, and the probability that an infectious needle is rendered virus free by an uninfectious addict is independent of the probability of transmission of HIV from the needle to the susceptible addict. The former assumption is very difficult to either verify or refute due to the lack of data on how addicts and needles interact with each other. The latter assumption is also open to question given that if an infectious needle is flushed (rendered virus free) by an initially uninfectious addict then all the infectious contents must enter the bloodstream of the addict which presumably increases the probability of HIV infection. Greenhalgh and Hay (1997) extend Kaplan's original model to allow infectious addicts to leave a needle virus free after use and incorporated a joint probability distribution between the transmission probability of HIV and the probability that a needle is flushed. In addition they also allowed for the possibility that addicts who discover that they are HIV positive stop or at least reduce their level of needle sharing. Greenhalgh and Hay found that assuming a joint probability distribution between α, the probability that a susceptible addict is infected with HIV in a single injection with an infectious needle and θ, the probability that a needle is flushed, made no difference to the long term prevalence of disease in the population. The effect of allowing infectious addicts to leave a needle virus free after use was difficult to assess due to a lack of data. They also found that if addicts who discover that they are HIV positive reduce their rate of sharing needles, then the long term prevalence of disease could be significantly reduced. This result suggests that introducing HIV testing for intravenous drug addicts could be potentially important as a control strategy, for example in conjunction with a needle exchange program. Kretzschmar and Wiessing (1998) also examine the effect of testing addicts for HIV and find that it is of very little benefit. However this result could depend heavily on the assumptions made by Kretzschmar and Wiessing relating to the infectivity of addicts during their infectious lifetime. We discuss this paper in detail later in this section.

So far we have discussed Kaplan's basic model and realistic extensions to this model.

We now discuss other work not directly based on Kaplan's models. We have noted that heterogeneous mixing in addicts is both more realistic and gives long term prevalence results which differ from homogeneous models. Capasso et al. (1995) discuss a deterministic model which assumes that addicts share needles in "friendship groups", these are as the name suggests groups of social acquaintances. They first deal with a single group model derived from the basic single population SIR model. They show that for the disease to become endemic in the population the basic reproductive number must exceed unity. If the basic reproductive number is less than or equal to unity then the disease will die out in all addicts and all needles. Later in Capasso et al. (1995) the single group model is extended to cater for the multi-group case.

Gani and Yakowitz (1993) use a random allocation model to examine the spread of HIV by needle sharing amongst small groups of intravenous drug users who are friends or relatives (buddy-users). They use a Markov chain approach to track the increase in the number of infectious drug users among stable groups of addicts. This model is used to examine the effects of the probability that addicts share needles and the probability of HIV transmission from an infectious needle to an addict in a single injection. Gani and Yakowitz later incorporate the replacement of infectious addicts from the group by uninfectious addicts from the general intravenous drug injecting population.

Yakowitz (1994) explores moving from a "microcosmic IVDU model to a macrocosmic HIV epidemic", (where IVDU stands for intravenous drug user). He uses a stochastic simulation approach to model the transmission of HIV among a population of drug users who meet on a periodic basis to share needles and inject drugs. This is the microcosmic model (which has similarities to the model discussed in Gani and Yakowitz (1993)). Yakowitz then extends this situation to allow for the possibility that IVDUs might circulate within an extended population, and be replaced by randomly selected individuals from that population. This represents the move from a microcosmic model to a macrocosmic model. He finds that this circulation model can exhibit a variety of possible behaviours and considers that due to the appreciable variability in the model it is not possible to find adequate deterministic approximations. Yakowitz finds somewhat unusually that in his model the prevalence of HIV does not reach a quasi-equilibrium steady state, this is contrary to the general behaviour of HIV/AIDS models which usually tend to a globally stable endemic equilibrium (Blower and Dowlatabadi, 1994).

Allard (1990) discusses a mathematical model which describes the risk of infection

from sharing injection equipment. He uses a probabilistic (as opposed to dynamic) model which examines the risk of infection from HIV each time that an addict injects with a shared needle. This model treats the risk of infection of HIV as a function of the following variables: the number of needles shared by the current addict; the probability that a needle is left infectious after use by an infectious addict; the probability of transmission of HIV from needle to addict; the number of addicts who have previously used each shared needle; the prevalence of HIV in the group from which the previous users are drawn; the number of times each needle has been used by each previous user and finally the number of times the current addict uses the needle. Allard claims that this model suggests that when each needle has been used previously by only one addict, the number of addicts with whom needles are shared is more important than the number of needles shared, and that the reverse is true when a needle has been used previously by many people.

Rather than construct a model of the spread of HIV via needle sharing from first principles, an interesting and possibly useful analogy can be made between the spread of HIV via sharing infected needles and the spread of malaria via infected mosquitoes biting humans. Massad et al. (1994) explore this connection and develop a new approach for the estimation of the basic reproductive number for HIV among IVDUs. This approach is based on an adaptation of the models proposed by Ross (1915, 1916) and Macdonald (1950, 1952, 1953) for vector-borne infections. However there are several obvious differences between the behaviour of needles and the behaviour of mosquitoes, for example only a fraction of infected mosquitoes are considered to be infective and mosquitoes have an incubation period before becoming infective. In contrast it is assumed that all needles which are infected are infective and there is no intrinsic incubation period of HIV in needles. Massad et al. adjust their model to take account of these behavioural differences so as to reflect how HIV is spread amongst a homogeneous addict population. One disadvantage of using this mosquito analogy is that whilst the model can be adjusted to fit the behaviour of needles, the interpretation of the parameters in this model is not as clear as in Kaplan's original model. For example it is convenient to be able to explicitly feature the probability that a needle is cleaned prior to use or flushed during use. Later in Massad et al. (1994) the model is extended to cover the possibility that a needle is not left infectious after use by an infectious addict and allows for addicts to inject at different rates. They also look briefly at interactions between distinct communities of IVDUs.

Blower et al. (1991) use a data-based mathematical model to assess the epidemiological consequences of heterosexual, intravenous drug use and perinatal transmission in New York City, USA. This is a complex model which consists of 34 ordinary differential equations and a large number of behavioural parameters. They use a scenario analysis of this model to examine the relationship between heterosexual and intravenous drug use transmission and provide a qualitative and quantitative insight into the HIV epidemic in New York City. They find that the behaviour of IVDUs has important knock-on effects for the heterosexual transmission of HIV. The model was used to predict the future number of adult and pediatric AIDS cases, however due to uncertainties in the parameter estimates in the model the confidence intervals of these predictions were very wide. An interesting result of this work is that Blower et al. claim that of the thirty or more variables in their model only a few of these were significant in contributing to the AIDS case prediction variability. The model suggests that accurate long term estimates of future numbers of AIDS cases will only be possible once accurate estimates of the key parameters are available.

Blower and Dowlatabadi (1994) use the complex deterministic model in Blower et al. (1991) as an example of carrying out a sensitivity and uncertainty analysis in a complex model of disease transmission. Uncertainty analysis may be used to assess the variability (prediction imprecision) in the outcome variable that is due to the uncertainty in estimating the input values. A sensitivity analysis can extend an uncertainty analysis by identifying which parameters are important in contributing to the prediction imprecision. Blower and Dowlatabadi travel around the sample space of the input parameters of their model using a Latin Hypercube sampling scheme. This is an extreme version of a Latin Square experimental design. Once the input parameters have been chosen the model in Blower et al. (1991) is simulated and the partial rank correlation coefficient (PRCC) between each input parameter and the output parameter (the value we wish to predict) is estimated. The size of the PRCC gives an indication of the relative importance of each parameter in the model in affecting the outcome variable. The sensitivity analysis suggests, perhaps not surprisingly, that the HIV transmission probability for intravenous drug users is one of the most influential model parameters. An interesting theoretical feature of this paper is the decision to use a Latin Hypercube sampling scheme as this design assumes that the effect of each input parameter is independent of the values of all other model parameters. This seems rather unlikely in a complex non-linear model.

All the articles we have discussed so far have explored the effects of behavioural changes on the spread of HIV/AIDS, such as the rate at which addicts share needles, the frequency at which needles are cleaned prior to use and the different social groups in which addicts share needles. Whilst these effects are important (and the need for more work in this area is reflected in the current MAP provisional report recommendations) they ignore the potential importance of epidemiological effects such as the three stages of infectivity which addicts progress through during the AIDS incubation period. As noted in Section 1.8 this is an area which is suggested by Anderson and May (1991) as important in studying the dynamics of HIV infection. However work in this area is very sparse with many researchers simply ignoring this effect (Seitz, 1998, personal communication). Several articles which do explore the effect of variable infectivity over the lifetime of drug users are Peterson et al. (1990), Seitz and Müller (1994), Tan and Tang (1993) and Kretzschmar and Wiessing (1998).

Peterson et al. (1990) use a complex Monte Carlo simulation model to investigate both behavioural and epidemiological effects of HIV infection in populations of intravenous drug users. Their model consists of three interacting sub-models: a model of HIV disease progression within an infected individual; a model describing the heterogeneity of intravenous drug use within needle sharing injecting communities; and a model of the social networks describing the pattern of needle sharing in drug addicts.

The stochastic model of HIV progression is based on the work of Longini et al. (1989) who used a four state Markov model to estimate the distribution of duration in each phase of HIV infection. The four stages in this model were Acute Infection, Asymptomatic, Pre-AIDS Symptoms and AIDS. This is in agreement with the assertion of three infectious stages during the incubation period of AIDS by Anderson and May (1991). Using data on viral antigen recovery and epidemiological data from blood transfusion recipients, (Ward et al., 1987), Peterson et al. estimate that the relative infectivity of addicts in the stages: Acute Infection: Asymptomatic: Pre-AIDS Symptoms is approximately 5:1:3. The second sub-model deals with the different sharing rates of addicts. This sub-model assumes that the lifecycle of an intravenous drug user starts from recruitment from the non-drug using population at large and then progresses sequentially into monthly, weekly and daily drug use. The final sub-model examines the social network that describes the relationships of people with whom a drug addict shares needles. This sub-model has three distinct types of needle sharing experiences: acquaintance sharing in small groups, random sharing with individual strangers, and

pooled sharing, resulting from the use of common injection equipment in a shooting gallery setting.

Peterson et al. put these three sub-models together to create a stochastic simulation model of how HIV spreads among a population of intravenous drug users. Using a variety of parameter estimates and simulations they report several interesting findings. Firstly they suggest that it is imperative that prevalence studies of HIV accurately report which type of drug addict is being surveyed: a heavy user or an occasional user. The model used by Peterson et al. suggests that the difference in prevalence between these two user groups can be very large indeed. Hence to get a balanced estimate of the prevalence of HIV in any given drug addict population care is required in differentiating between these two types of users. Peterson et al. also find that the initial "seeding" in their model has a large effect on the confidence interval of long term prevalence in the population. In other words how the disease initially progresses seems to play an important part in the long term level of disease in the population.

Whilst Peterson et al. incorporate a three stage infectious period in HIV positive addicts they do not directly compare the effect of moving from a constant infectious period to three stage infectivity. This effect is explored by Seitz and Müller (1994) who model the spread of HIV in the population at large (including drug addicts, heterosexual and homosexual population groups). Seitz and Müller assume that the infectivity in an HIV positive individual has a so-called "bath-tub" shape. They justify this by arguing that at the onset of infection, during the first several weeks before the body develops an immune response, the virus is free to replicate unabated. In advanced states of the disease, the immune system becomes increasingly impaired and viral replication again accelerates. Both phases bear directly on HIV transmission because HIV infectivity is thought to be related to the viral load in the infected individuals bodily fluids, (Osmond, 1990). Hence the "bath-tub" effect represents the higher infectivity of an HIV positive individual at the start of the AIDS incubation period and during the final full blown AIDS stage.

Seitz and Müller use a deterministic simulation model to examine the effects of moving from the conventional assumption of constant infectivity to three stage infectivity. The model used is highly complex and uses many biological and behavioural parameters. A particularly interesting feature of this model is that the incubation period of AIDS is increased once the simulations reach 1987 to reflect better drug therapy. One perhaps dubious assumption in this model is that the "bath-tub" infectivity assumes

that the second rise in infectivity occurs at the onset of full blown AIDS (also called frank AIDS). This seems sensible in terms of the infectivity of an individual, however AIDS epidemic modellers usually assume that people with frank AIDS, who are usually aware of their condition, do not continue to share needles or engage in unprotected sexual activity (Kaplan, 1995, Kaplan, 1989a, Anderson and May, 1991).

Seitz and Müller incorporate three stages of infectivity by assigning relative infectivity factors to individuals according to how much time has elapsed since the initial infection with HIV. It is important to note that these infectivity factors do not only adjust for the differing viral load in infected persons but also their lifestyle and behaviour. For example it is assumed that the infectivity of heterosexual men is 169 times greater after initial infection with HIV than in the later Asymptomatic stage, similarly at the onset of frank AIDS the infectivity is 25 times greater than in the Asymptomatic stage. The infectivity factor of 25 at the onset of frank AIDS is justified as a combination of increased viral load but decreased sexual activity through choice and reason of illness.

Seitz and Müller simulate their model for a number of different scenarios (parameter choices) and compare the results with the case where the infectivity of an individual is constant during their infectious lifetime, and importantly where the cumulative infectivity over the infectious lifetime is the same as in the three stage infectivity case. The results of these comparisons are interesting in that three stage infectivity resulted in a greatly increased incidence of HIV and AIDS in the total population. In addition it appears that the dynamics of the epidemic is different under three stage infectivity with a large wave of HIV incidence occurring at the start of the epidemic which dies out as time progresses. In the constant infectivity case the disease progresses much more slowly among the population. This work by Seitz and Müller suggests that moving from constant (single stage) stage infectivity to three stage infectivity could have a large effect on the long term prevalence of HIV and AIDS in the population. A slight criticism of this work is that there is a lack of transparency in the model used by Seitz and Müller (which is probably due to the considerable complexities of the model). It would have been interesting to examine more closely the effects of three stage infectivity in individual population groups (such as intravenous drug users) to assess more clearly why such a large increase in long term prevalence was observed and what particular assumptions were made relating to the behaviour of each population group.

Tan and Tang (1993) develop a stochastic model for the HIV epidemic involving both sexual contact and intravenous drug use. Their model is formulated in terms

of a chain multi-nomial model which is used to derive the expected sizes of different population classes throughout the course of an HIV epidemic. Tan and Tang divide the population up according to whether individuals are susceptible, infectious with HIV or have developed full blown AIDS, in addition a class is reserved for individuals who have been infected but are not yet infectious. The class of infectious individuals is also partitioned into k sub-stages to allow for varying levels of infectivity during the infectious lifetime. The model also contains separate subgroups according to the sexual and drug use behaviour of individuals. Tan and Tang initially consider a discrete time stochastic model before moving on to the analogous continuous time situation. An interesting aspect of this work is that the expected numbers of susceptible, latent, infectious and AIDS infected individuals in the population under the stochastic model can differ markedly from the deterministic equivalent of this model. The paper concludes with a simulation study of the model using a variety of parameter estimates. The relative infectivity of infectious individuals and the duration of each infectious stage was determined by using data on clinical observations of HIV infection (Redfield and Burke, 1988). Tan and Tang chose a latent period of 1.25 months duration followed by a five stage infectious period where these stages have duration 7 months, 4.17 years, 1.67 years, 1.67 years and 1.67 years respectively. It was assumed that the infectivity of an individual monotonically increases with time during the first infectious stage reaching a maximum infectivity after four months, after which the infectivity of an individual decreases to a very low level prior to entering the second infectious stage. Infectivity remains very low during stage two before again increasing monotonically throughout stages three, four and five towards the AIDS infectivity stage.

Kretzschmar and Wiessing (1998) examine the spread of HIV among a population of drug users using a stochastic simulation model. The main aspects of addict behaviour which they examine are the social networks in which addicts share needles and the frequency with which needles are shared. Kretzschmar and Wiessing construct their model by examining the behaviour of two different populations of addicts in Holland. The model used distinguishes between two different types of needle sharing: "buddy" sharing and "stranger" sharing. When addicts share needles with a sexual partner or friend this is classed as "buddy" sharing, whereas when addicts visit shooting galleries this is classed as "stranger" sharing. Kretzschmar and Wiessing claim that the majority of sharing in Holland is "buddy" sharing but there does exist a small core of high risk users (those who have a high sharing rate with "strangers"). Their simulation

model allows addicts to join and leave "buddy" relationships over the course of their sharing lifetime and allows addicts to share needles at different rates in the different relationships. In addition to modelling the type of needle sharing undertaken by addicts the model also incorporates variability in the infectivity of addicts who are infected with HIV. They assume that an infectious addict progresses through two distinct phases of infectivity until the addict develops full blown AIDS (at which stage they are removed from the sharing population). Kretzschmar and Wiessing assume that during the first 60 days after initial infection the probability of transmission is 0.5 per contact (in other words sharing a needle with an addict in this stage carries a 50% risk of infection). After 60 days the infectivity of an infectious addict then drops to 0.001 per contact. The mean sojourn time in the population is taken to be ten years. They ensure that the mean infectivity of an infectious addict in their model is the same as that estimated by Kaplan and Heimer (1992a), namely 0.01 per contact. It is argued that the state of health of an addict once the viral load starts to increase prior to the development of full blown AIDS is such that sharing does not occur, hence a two stage infectious period is used rather than a three stage infectious period.

Kretzschmar and Wiessing use their model to examine a number of prevention strategies. The strategies used consist of lowering the frequency at which various groups of addicts share needles, in particular several strategies examine the effect of testing addicts for HIV. If an addict tested positive then his or her frequency of sharing was reduced. After studying the output from their model they find that reducing sharing with "stranger" users is more effective than reducing the overall frequency of sharing. They also find that there is a threshold sharing frequency below which the epidemic never takes off. A particularly interesting finding is that the model suggests that performing an HIV test on 10% of addicts or even 50% of addicts a year has no appreciable effect on HIV incidence (a positive HIV test results in the future frequency of sharing being reduced by a factor of 0.5). They find that the only effective prevention strategies are those that reduce sharing frequencies in the entire population. It should be noted that, as acknowledged by Kretzschmar and Wiessing, these results depend heavily on the infectivity assumptions in their model, in particular the very low infectivity of an addict after the first 60 days of infection. For example as far as the HIV testing strategy is concerned, a reason for this strategy being ineffective is that by the time most addicts are tested they have already reached the low infectivity period and as such are of limited importance in spreading the epidemic compared to newly infectious addicts.

A smaller difference in the infectivity of addicts in the two infectious stages (or if the model included a third infectious stage as in Peterson et al. (1990)) may make HIV testing a more attractive prevention strategy.

Hay (1999) examines a number of mathematical models of both a deterministic and stochastic nature. He first discusses a deterministic model due to Kaplan (1989a) and then converts this into a stochastic model. This stochastic model allows for injection equipment to have a varying (time dependent) level of infectivity. Hay examines the effect of moving from the assumption that injection equipment has a constant infectivity to three different infectivity assumptions. Firstly the infectivity of injection equipment is assumed to decrease over time according to an exponential distribution, then it is assumed that infectivity decreases linearly and finally that injection equipment remains infectious for a fixed number of days after which infectivity drops to zero. In addition Hay also examines the case where infectivity is split into three distinct stages in a similar fashion to Seitz and Müller (1994). He finds that assumptions relating to the infectivity of injection equipment have a significant effect on the dynamic behaviour of an HIV epidemic.

1.11 A Basic Needle Sharing Model

All the models we discuss in this thesis are extensions of the basic needle sharing model due to Kaplan and O'Keefe (1993). The modelling assumptions in this basic model are discussed in detail in the following chapter. For the moment we simply give a brief description of this model and state its equations together with some comment as to its main properties.

Firstly we assume that the population amongst whom the disease is spreading is of size n, where n is large; the random variability in the fraction of infected addicts and needles at time t is sufficiently small to be ignored; and that addicts who leave the population for any reason are immediately replaced by susceptible addicts. We also assume that all addicts and needles mix randomly and behave according to independent Poisson processes. We suppose that addicts inject with shared needles at rate λ per unit time, leave the population for non disease related reasons at rate μ per unit time and that infected addicts develop AIDS at rate δ per unit time (at which point they leave the sharing, injecting population). We also assume that a needle is always left infectious after use by an infectious addict and the probability of HIV transmission from an

infectious needle to a susceptible addict in a single injection is α. In addition we suppose that a susceptible addict can flush (render uninfectious) an infectious needle during the injection process with probability θ. Finally addicts successfully clean needles prior to use with probability ϕ and each needle in the population is exchanged for an uninfectious needle at rate τ per unit time.

Using these assumptions and letting $\pi(t)$ and $\beta(t)$ respectively denote the fraction of addicts and needles that are infected with HIV at time t, and defining the ratio of addicts to needles by $\gamma = n/m$, the following differential equations describe the spread of the disease:

$$\frac{d\pi}{dt} = (1-\pi)\lambda\beta\alpha(1-\phi) - \pi(\mu+\delta), \tag{1.1}$$

and

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\pi - \beta\lambda\gamma(1-\pi)(1-(1-\theta)(1-\phi)) - \beta\tau. \tag{1.2}$$

Equation (1.1) states that the prevalence of infectious addicts will increase when an uninfectious addict injects with an infectious needle which is not cleaned prior to use and HIV transmission occurs, and decrease when an addict develops full blown AIDS or leaves the sharing, injecting population for other reasons. Equation (1.2) states that the prevalence of infectious needles will increase when an infectious addict injects with a previously uninfectious needle, and decrease when a previously infectious needle is used by an uninfectious addict and is either cleaned or flushed during injection, or the needle is exchanged.

An endemic solution is possible in this model if and only if the parameter R_0 exceeds unity, where

$$R_0 = \frac{\lambda\alpha(1-\phi)}{(\mu+\delta)(\hat{\theta}+\hat{\tau})}. \tag{1.3}$$

Here $\hat{\theta} = 1-(1-\theta)(1-\phi)$ and $\hat{\tau} = \tau/(\lambda\gamma)$. As usual, R_0 has a natural biological interpretation: it is the total expected number of secondary infections caused by a single infectious addict during his or her entire infectious lifetime, after entering a population of uninfectious needles and otherwise susceptible addicts. In the following chapters we extend the model in eqns (1.1)-(1.2) to more realistically reflect the spread of HIV via needle sharing.

1.12 Summary and Comments

From our previous review we can draw several important points. Firstly it is apparent that infection from HIV is a serious global problem. However the actual method of infection varies around the world, the disease is mainly spread via heterosexual contact in the developing countries in sub-Saharan Africa, and mainly through intravenous drug use and homosexual contact in the developed countries of Western Europe and North America.

The discovery of AIDS in the early 1980's was followed by a large amount of research into both the biological nature of the disease and epidemiological modelling aspects. However twenty years on there is comparatively little interest in HIV and AIDS as in the developed world the disease has failed to take hold among the heterosexual population at large, contrary to some of the early doomsday predictions. Indeed HIV and AIDS is in some ways unique in that in the developed world it has so far only really affected minorities such as homosexual men and intravenous drug addicts. However while high risk populations such as intravenous drug users are comparatively small in number, it seems likely that if HIV was to begin to take hold among the population at large then it would stem from one of these minority groups. This is particularly true in some areas of post-communist Europe where the number of people participating in intravenous drug use has increased dramatically. Therefore an in-depth study of the transmission of HIV among populations of intravenous drug users seems a potentially important area of research.

We have previously discussed a number of articles which either deal solely with the spread of HIV through the sharing of contaminated injection equipment, or where this is contained in part of the model. There are a number of important heterogeneities involved in modelling the spread of HIV via needle sharing. For example addicts exhibit a wide range of needle sharing rates and the efficiency of needle cleansing is highly variable. This is a common feature examined in many of the articles previously discussed. Much less common is the study of how addicts and needles interact, for example examining the consequences of whether infectious addicts always leave needles infectious after use and uninfectious addicts always leave previously infectious needles uninfectious after use, and the impact the latter has on the probability of HIV transmission. As discussed by Greenhalgh and Hay (1997) these interaction assumptions play an important part in the disease dynamics.

As previously mentioned medical evidence suggests that the infectivity of an infectious addict may vary substantially throughout his or her infectious lifetime. This again is a feature not commonly examined in the literature and moreover those articles which do examine this use a wide variety of different assumptions. In particular we are unaware of any articles which incorporate addicts of different levels of infectivity and also make any reference to how these respective addicts interact with either uninfectious needles or needles of different levels of infectivity. It is difficult to see how one could adequately examine the effect of a three stage infectious period (for example) while not including these possibly crucial interaction assumptions. This brings us to the focus of the work in this thesis.

1.13 Thesis Overview

A main aim of this thesis is to establish whether the inclusion of a three stage infectious period into existing models of the spread of HIV via needle sharing affects their behaviour. For example we wish to determine whether moving from a single stage to a three stage infectious period increases the long term level of disease among the population. Alternatively we could ask whether the extra complexity required to model a three stage infectious period produces sufficiently different behaviour to warrant inclusion into a parsimonious model of HIV transmission via needle sharing. A related issue which we examine is the effectiveness of testing addicts for HIV as a method of reducing the spread of disease. We expect that an addict who knows he or she is infectious will reduce their level of needle sharing. Intuitively this control strategy should be directly affected by assumptions made relating to the infectivity of addicts. For example this strategy may be less effective if addicts are highly infectious after initial infection and then their infectivity decreases until their removal from the population, than if their infectivity continually increases during the AIDS incubation period.

There are several common themes throughout this thesis. First and foremost we try to keep the work practically focused since the motivation for studying models of disease transmission is to provide insight and practical guidance into the management of infectious diseases. Secondly whenever possible we try to adopt an analytical approach supported by numerical or simulation methods. The main motivation for this is that in the models we study there is generally a lack of good quality data from which to estimate some of the model parameters. This is particularly true when trying to estimate how

addicts and needles interact since we are unaware of any data to assist with this.

In Chapter 2 we discuss and extend the model due to Kaplan and O'Keefe (1993) to allow addicts to progress through three stages of infectivity. We refer to this model as the Simple Model. We conduct a stability analysis of this model and a short simulation study to validate our analytical results. The Simple Model assumes that while addicts have three different classes of infectivity, needles are still only classed as either infectious or not infectious.

In Chapter 3 we extend the Simple Model to also allow needles to exist in three infectious classes where each class corresponds to a class of infectious addict. In order to split the class of infectious needles into three infectious sub-classes it is necessary to make assumptions relating to how addicts and needles in the various infectious classes interact with each other. These assumptions are referred to as addict-needle interaction assumptions. In extending the Simple Model it was decided first to assume that any needle (whether infectious or not) adopts the infectivity characteristics of the current user. This is a rather extreme addict-needle interaction assumption and this model represents a lower bound of the spread of disease under three stage infectivity. We refer to this model as the Optimistic Model. We conduct a stability analysis of this model and a short simulation study to validate our analytical results.

In Chapter 4 we extend the Simple Model to again allow needles to exist in three infectious sub-classes where each class corresponds to a class of infectious addict. This time we extend the model by assuming that a needle always adopts the more infectious class between that of the needle prior to use and that of the current user. As in the Optimistic Model this is a rather extreme addict-needle interaction assumption but in this case our model represents an upper bound of the spread of disease under three stage infectivity. We refer to this model as the Pessimistic Model. We again conduct a stability analysis of this model and a short simulation study to validate our analytical results.

In Chapter 5 we develop a generalisation of the extreme Optimistic and Pessimistic Models which uses a general probability structure to define the outcome of each addict-needle interaction. This model is referred to as the General Model. We then show a number of analytical results relating to the behaviour of this model and a short simulation study to explore the behaviour which could not be shown analytically. In Chapter 6 we discuss extending the Simple Model and the General Model to incorporate

mortality due to AIDS and the recruitment of new (susceptible) drug users from the population at large. In previous chapters we have assumed instead that the drug addict and needle populations maintain a constant size. Several local stability results are shown followed by a short discussion on the effect of AIDS mortality and a number of simulations showing both the effect on the population size of an HIV/AIDS epidemic and the long term prevalence level of HIV in addicts.

In Chapter 7 we discuss the practical implications of allowing addicts and needles to exist in three states of infectivity. In short we explore whether our three stage models can offer new insight into the spread of HIV via needle sharing among intravenous drug users. This chapter compares the long term prevalence of HIV in the three stage models with single stage equivalents (such as the model used in Kaplan and O'Keefe (1993)). We are specifically interested in determining whether the long term prevalence of HIV is higher as a result of three stage infectivity (as suggested by Seitz and Müller (1994)). In addition the effect of control strategies such as needle exchange programs in the three stage models is examined along with the effect of improved needle cleaning.

In Chapter 8 we move away from investigating the effect of a three stage infectious period and examine the effect of testing addicts for the presence of HIV. We first extend the original model due to Kaplan and O'Keefe (1993) and then perform a stability analysis on this model. We then compare the effect of HIV testing in this model with that of a simpler model discussed by Greenhalgh and Hay (1997). We finally extend the Optimistic and Pessimistic Models to incorporate HIV testing and examine the impact that different relative infectivity assumptions have on the effectiveness of this control measure.

In Chapter 9 we discuss a method of sensitivity analysis suggested by Blower and Dowlatabadi (1994). Using the HIV test model from Chapter 8 as an example, we show that the sampling scheme used in this paper can produce misleading results. In Chapter 10 we examine stochastic equivalents to the deterministic models discussed in previous chapters and use simulation to ascertain whether the long term behaviour of these models is comparable. We also briefly look at a stochastic threshold theorem for the Kaplan and O'Keefe model. The thesis concludes with Chapter 11 containing a discussion and summary of the work in Chapters 2-10. We also discuss several suggestions for future work.

Chapter 2

The Simple Model

2.1 Introduction

In this chapter we develop and investigate the basic model that we shall use to describe the spread of HIV among a population of intravenous drug users. We first describe in detail the modelling assumptions used by Kaplan and O'Keefe (1993), and extend these to allow addicts to progress through three different stages of infectivity prior to the onset of AIDS. We then derive a system of differential equations based on this extended set of assumptions. Next we derive an expression for the basic reproductive number, R_0, and move on to investigating the behaviour of our model. In particular we are interested in the conditions necessary for the disease to die out or to persist in the population. We finally examine numerical simulations of our model in order to validate our previous mathematical results before concluding the chapter with a short summary of the main findings.

2.2 Kaplan and O'Keefe Model

Kaplan and O'Keefe (1993) describe a model which is itself an extension of a model due to Kaplan (1989a). The model featured in the later paper is significant in that it incorporates a needle exchange program. Such programs have been demonstrated to be an important measure in reducing the spread of HIV among intravenous drug users. Greenhalgh and Hay (1997) discuss in detail the model due to Kaplan (1989a). Kaplan describes a deterministic model for the spread of HIV amongst intravenous drug users, and it was assumed that the population amongst whom the disease is spreading is of size n, where n is large. He makes the following assumptions:

1. All sharing of drug injecting equipment occurs in shooting galleries. In the model a shooting gallery is defined as a location where addicts sequentially rent the same drug-injection equipment. There are m shooting galleries (or equivalently m "kits" of drug-injection equipment are in circulation) and addicts select shooting galleries (or "kits") at random. All addicts inject once per visit to a shooting gallery.

2. Each addict visits shooting galleries in accordance with a Poisson process with rate λ, independently of the actions of other addicts.

3. Injection equipment always becomes infectious if it is used by an infected addict. When infectious injection equipment is used by an uninfected addict the act of injecting will replace the infectious blood in the needle with uninfectious blood from the addict with probability θ. When this occurs the needle is said to have been "flushed". Any uninfected addict who uses infectious injection equipment is considered to be exposed to HIV.

4. Given exposure to HIV an addict becomes infected with probability α; α is the infectivity of HIV via shared injection equipment. Sharing injection equipment is the only means by which addicts may become infected.

5. Infectious addicts develop full blown AIDS according to a Poisson process with rate δ, at this stage addicts leave the sharing, injecting population. These addicts are immediately replaced by susceptible addicts.

6. Infectious addicts depart the population for reasons other than developing full blown AIDS (for example due to death, treatment with methadone, or relocation) at rate μ and are immediately replaced by susceptible addicts.

7. The random variability in the fraction of infected addicts and needles at time t is sufficiently small to be ignored.

The Kaplan and O'Keefe extension to Kaplan (1989a) additionally assumes that:

8. An addict effectively cleans (or bleaches) the injection equipment immediately prior to use with probability ϕ.

9. Each needle is exchanged (or renewed) for an uninfected needle according to a Poisson process with rate τ.

As previously discussed it has been demonstrated by Peterson et al. (1990) and Anderson and May (1991) that once an individual is infected with HIV, the viral load of HIV in the blood varies considerably over the lifetime of the individual. The viral load can be interpreted as the amount of HIV virus per unit volume of blood, hence viral load is a measure of the infectivity of an individual. The variability in viral load can be approximated into three distinct sequential stages: stage one - Acute Infection, stage two - Asymptomatic and stage three - Pre-AIDS symptoms. The Pre-AIDS stage is where the addict is classed as having AIDS Related Complex (ARC), following this stage an addict develops full blown AIDS (at which point they leave the needle sharing, injecting population). We now incorporate the Acute Infection, Asymptomatic and Pre-AIDS stages into Kaplan and O'Keefe's model by replacing Assumption 5 with the following model assumptions:

5a. Immediately after the initial infection an addict is defined to be Acutely Infectious and enters the Asymptomatic stage according to a Poisson process with rate δ_1.

5b. Asymptomatic addicts enter the Pre-AIDS stage according to a Poisson process with rate δ_2.

5c. Pre-AIDS addicts enter the full blown AIDS stage according to a Poisson process with rate δ_3, at this point addicts leave the sharing, injecting population.

2.3 Model Derivation

We now derive the differential equations which define the spread of HIV among an intravenous drug addict population where addicts progress through three stages of infectivity prior to the onset of AIDS. We derive four equations, one for each stage of infectious addict and one for infectious needles.

The number of stage one infected addicts at time $t + \Delta t$

$$
\begin{aligned}
= \quad & \{\text{number of stage one addicts at time } t\} \\
& + \{(\text{number of uninfected addicts at time } t) \\
& \times (\text{fraction of addicts who inject in } [t, t+\Delta t) \text{ with an infectious needle which is not cleaned prior to use and where transmission of HIV occurs in a single injection})\}
\end{aligned}
$$

$-\{$number of stage one infected addicts who progress into stage two infectivity or leave the sharing, injecting population in $[t, t+\Delta t)\}$.

Thus

$$\begin{aligned} n\pi_1(t+\Delta t) &= n\pi_1(t) + n(1-\pi_1(t)-\pi_2(t)-\pi_3(t))\lambda\Delta t\beta(t)\alpha(1-\phi) \\ &\quad -n\pi_1(t)\Delta t(\mu+\delta_1) + o(\Delta t). \end{aligned}$$

Subtracting $n\pi_1(t)$ from both sides, dividing by $n\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\pi_1}{dt} = \Big(1 - \sum_{i=1}^{3} \pi_i\Big)\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\pi_1.$$

The number of stage two infected addicts at time $t+\Delta t$

$=$ $\{$number of stage two addicts at time $t\}$

$+\{$number of stage one addicts who enter the stage two infectious class in $[t, t+\Delta t)\}$

$-\{$number of stage two addicts who enter the stage three infectious class or leave the sharing, injecting population in $[t, t+\Delta t)\}$.

Thus

$$n\pi_2(t+\Delta t) = n\pi_2(t) + n\pi_1(t)\delta_1\Delta t - n\pi_2(t)(\mu+\delta_2)\Delta t + o(\Delta t).$$

Subtracting $n\pi_2(t)$ from both sides, dividing by $n\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2.$$

Similarly

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3.$$

The number of infected needles at time $t+\Delta t$

$=$ $\{$number of infected needles at time $t\}$

$+\{$(number of uninfected needles at time t) $\times$ (fraction of needles used by infected addicts in $[t, t+\Delta t))\}$

$-\{$(number of infected needles at time t) $\times$ (fraction

of infected needles used by uninfectious addicts in $[t, t+\Delta t)$

and left in an uninfected state)}

$-\{$number of infected needles exchanged in $[t, t+\Delta t)\}$.

Thus

$$\begin{aligned} m\beta(t+\Delta t) &= m\beta(t) + m(1-\beta(t))\lambda\Delta t\gamma(\pi_1(t)+\pi_2(t)+\pi_3(t)) \\ &\quad -m\beta(t)\lambda\Delta t\gamma(1-\pi_1(t)-\pi_2(t)-\pi_3(t))(1-(1-\phi)(1-\theta)) \\ &\quad -m\beta(t)\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m\beta(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - \beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - \beta\tau.$$

Hence the system of differential equations which describes the spread of the disease is:

$$\frac{d\pi_1}{dt} = \Big(1-\sum_{i=1}^{3}\pi_i\Big)\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\pi_1, \tag{2.1}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2, \tag{2.2}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3, \tag{2.3}$$

and

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - \beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - \beta\tau, \tag{2.4}$$

with suitable initial conditions: $0 \le \pi_1(0), \pi_2(0), \pi_3(0), \beta(0)$ and $\pi_1(0)+\pi_2(0)+\pi_3(0)$, $\beta(0) \le 1$.

2.4 The Basic Reproductive Number

The basic reproductive number is commonly defined as the expected number of secondary infections caused by a single newly infectious individual entering a totally susceptible population at equilibrium (Diekmann et al., 1990). This number is a function of the model parameters and is denoted by R_0. In many epidemiological models the value of R_0 is of fundamental importance, in particular $R_0 = 1$ is commonly a threshold value which when crossed causes radically different behaviour in the model concerned.

When $R_0 < 1$ we expect the disease to die out, whereas if $R_0 > 1$ we expect the disease to take off. We now derive an explicit expression for R_0 based on the model defined by equations (2.1)-(2.4).

Considering a single newly infected addict entering a population containing only susceptible addicts and uninfectious needles at the disease-free equilibrium, the initial infection process can be broken down into two distinct phases. Firstly the disease passes from our single infectious addict to an uninfectious needle, secondly this needle (which is now infectious) passes on the disease to a susceptible addict. We wish to find the expected number of needles a single addict will infect during his or her infectious lifetime and the expected number of addicts each of these needles will infect. The product of these expected values is R_0.

Addicts progress through three infectious stages. During each stage an addict will leave needles infectious. Addicts inject at rate λ per unit time and spend on average $1/(\mu + \delta_1)$ time units in stage one. An addict progresses from stage one to stage two with probability $\delta_1/(\mu + \delta_1)$ and spends on average $1/(\mu + \delta_2)$ time units in this stage. Similarly an addict progresses from stage two to stage three with probability $\delta_2/(\mu+\delta_2)$ and spends on average $1/(\mu + \delta_3)$ time units in this stage. Hence on average an addict infects

$$\frac{\lambda}{\mu + \delta_1} + \frac{\lambda\delta_1}{(\mu + \delta_1)(\mu + \delta_2)} + \frac{\lambda\delta_1\delta_2}{(\mu + \delta_1)(\mu + \delta_2)(\mu + \delta_3)}$$

needles during his or her entire infectious lifetime.

We now determine how many infections are caused by each needle until it is rendered virus free. Consider a single infectious needle, we want to find E(addicts infected by this single needle), the expected number of addicts infected by this needle. To find this value we first condition on the outcome of the next event, that of a needle being rendered virus free *before* the next user injects with it. We partition this event into two, either the needle is rendered virus free before the next injection or it is not. Let Y denote the number of addicts infected by a single needle, let X_1 denote the event that the needle is rendered safe before the next injection, and let X_2 denote the event that the needle is still infectious at next injection. Therefore we have that

$$E(Y) \;=\; E(Y \,|\, X_1)P(X_1) + E(Y \,|\, X_2)P(X_2).$$

If the needle is rendered safe prior to the next injection then the infected needle has infected zero addicts, thus $E(Y \mid X_1) = 0$. The event X_2 corresponds to the needle

being neither cleaned nor exchanged prior to use. The probability of this event is $\lambda\gamma(1-\phi)/(\lambda\gamma+\tau)$, hence

$$E(Y) = E(Y \mid X_2)\frac{\lambda\gamma(1-\phi)}{\lambda\gamma+\tau}.$$

We now explore $E(Y \mid X_2)$ by conditioning on the next event, that of a susceptible addict injecting with an infectious needle. This event has four outcomes. An addict may be infected by the needle or still remain susceptible, in addition the addict may leave the needle infectious or flush the needle during use. Independence of the events that an addict is infected by the needle and the needle is flushed is not necessary (Greenhalgh and Hay, 1997). Also it is not realistic either as if the needle is flushed the addict is more likely to be infected.

Consider the event that a susceptible addict injects with an infectious needle, each of the four outcomes mentioned previously are possible and each outcome has different implications for the number of addicts infected by this needle. Suppose that the addict flushes the needle, this means that the needle cannot infect any other addicts. If the addict becomes infected then the total number of addicts infected from the needle is one, if the addict is not infected then the total number of addicts infected from the needle is zero. Suppose that the addict does not flush the needle, this means that the needle is still infectious after the addict has used it and now awaits the next user. If the needle is not flushed the infection process has been renewed, the needle is in the same state as before but now has either one or zero infections to its credit depending upon whether the susceptible addict was infected or not. In the former case the needle will infect $E+1$ addicts where $E = E$(addicts infected by a single needle), similarly if the addict was not infected then the needle will infect E addicts. We can express this as

$$\begin{aligned} E &= \frac{(1-\phi)\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected and needle left infectious)}(1+E) \\ &\quad + \text{P(sus. addict infected and needle left virus free)} \\ &\quad + \text{P(sus. addict not infected and needle left infectious)}E\Big], \\ &= \frac{(1-\phi)\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected)} + \text{P(needle left infectious)}E\Big], \\ &= \frac{(1-\phi)\lambda\gamma}{\lambda\gamma+\tau}[\alpha+(1-\theta)E]. \end{aligned}$$

Solving for E gives

$$E = \frac{(1-\phi)\alpha}{\hat{\tau}+\hat{\theta}},$$

where $\hat{\tau} = \tau/\lambda\gamma$ and $\hat{\theta} = 1-(1-\phi)(1-\theta)$.

We now have the expected number of addicts infected by a single needle, multiplying this by the expected number of needles an addict infects during his or her entire infectious lifetime gives R_0. Hence,

$$R_0 = \frac{\lambda\alpha(1-\phi)}{(\mu+\delta_1)(\hat{\tau}+\hat{\theta})}\left[1+\frac{\delta_1}{\mu+\delta_2}+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right]. \tag{2.5}$$

We now move on to investigate the properties of our model and in particular we are interested in what role R_0 plays in determining long term behaviour.

2.5 Analytical Results

We now turn our attention to the behaviour of our model and use analytical results to illustrate key properties. We are primarily interested in the properties of the equilibrium solutions, in particular how many steady state solutions the model possesses and whether these solutions are stable or unstable. We place two restrictions on the values of the model parameters, firstly we assume that all parameters with the exception of ϕ, the probability that an addict successfully cleans a needle prior to use, are strictly positive; this is necessary to avoid complications such as dividing by zero. We allow ϕ to take on the value zero as this represents the practically important situation where addicts do not practice any kind of preventative cleaning prior to injecting with a shared needle, we also assume that ϕ is strictly less than unity. If $\phi = 1$ then HIV transmission is impossible and the model is meaningless (note also that $\phi = 1$ implies that $R_0 = 0$).

Define the region D in $\mathbf{R}^4$ by $D = [0,1]^4$. The system defined by differential equations (2.1)-(2.4) starts in the region D. The right-hand sides of these equations are differentiable with respect to π_1, π_2, π_3 and β, with continuous derivatives, and the corresponding vector points into the region D on its boundary except at the origin, which is clearly an equilibrium point. It is straightforward using standard techniques (Hale, 1969) that equations (2.1)-(2.4) with initial conditions in D, have a unique solution that remains in D for all time.

Theorem 2.1 *If $R_0 \leq 1$ the system of equations (2.1)-(2.4) has a unique equilibrium solution where the disease has died out in both addicts and needles. If $R_0 > 1$ then there*

is still the equilibrium where the disease has died out, however there is also a unique endemic equilibrium.

Proof.

Suppose that $\pi_1^\star$, $\pi_2^\star$, $\pi_3^\star$ and $\beta^\star$ denote respectively the equilibrium values of π_1, π_2, π_3 and β, and $\pi = \pi_1 + \pi_2 + \pi_3$. From equations (2.2) and (2.3) we have that $\pi_2^\star = (\delta_1\pi_1^\star)/(\mu+\delta_2)$ and $\pi_3^\star = (\delta_2\pi_2^\star)/(\mu+\delta_3)$, so $\pi^\star = \pi_1^\star + \pi_2^\star + \pi_3^\star = \pi_1^\star L$, where

$$L = 1 + \frac{\delta_1}{\mu+\delta_2} + \frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}.$$

From eqn (2.1) we find that

$$\beta^\star = \frac{\pi^\star(\mu+\delta_1)}{L(1-\pi^\star)\lambda\alpha(1-\phi)}. \tag{2.6}$$

From eqn (2.4) we find that

$$\beta^\star = \frac{\pi^\star}{\pi^\star + (1-\pi^\star)\hat{\theta} + \hat{\tau}}. \tag{2.7}$$

Equating (2.6) and (2.7) and dividing by $\pi^\star$ (assuming that $\pi^\star \neq 0$) gives us

$$\pi^\star = \frac{\lambda\alpha(1-\phi)L}{(\mu+\delta_1)(1-\hat{\theta}) + \lambda\alpha(1-\phi)L}\left(1 - \frac{(\mu+\delta_1)(\hat{\tau}+\hat{\theta})}{\lambda\alpha(1-\phi)L}\right). \tag{2.8}$$

We now know $\pi^\star$, $\beta^\star$ is found by substitution into (2.6), hence

$$\beta^\star = \frac{1}{1+\hat{\tau}}\left(1 - \frac{(\mu+\delta_1)(\hat{\tau}+\hat{\theta})}{\lambda\alpha(1-\phi)L}\right). \tag{2.9}$$

Using the expression for R_0 in eqn (2.5) we find that

$$(\pi^\star, \beta^\star) = \left(\frac{\lambda\alpha(1-\phi)L}{(\mu+\delta_1)(1-\hat{\theta}) + \lambda\alpha(1-\phi)L}\left(\frac{R_0-1}{R_0}\right), \frac{1}{1+\hat{\tau}}\left(\frac{R_0-1}{R_0}\right)\right). \tag{2.10}$$

It is obvious that there can only be two equilibrium solutions, the disease-free solution where $\pi^\star = 0$ and $\beta^\star = 0$ and a strictly positive solution. From eqn (2.10) we see that if $R_0 \leq 1$ then the only solution is the disease-free solution, if $R_0 > 1$ then we also have a unique positive solution. This completes the proof.•

Theorem 2.2 *If $R_0 \leq 1$ then whatever the initial state the disease will die out in both addicts and needles.*

Proof.

The key stage in the method of the proof is to show that $\lim_{t\to\infty}\pi_1(t) = 0$. We prove this result in several stages. Let $\bar{\pi}_1(t) = \sup_{\xi\geq t}\pi_1(\xi)$, this is monotone decreasing in t. For all t we have that $\pi_1(t) \leq \bar{\pi}_1(t)$. Hence, given $\epsilon > 0$ there exists $t_1(\epsilon)$ such that $\pi_1(t) \leq \pi_1^\infty + \epsilon$ for all $t \geq t_1(\epsilon)$ where $\pi_1^\infty = \limsup_{t\to\infty}\pi_1(t) = \lim_{t\to\infty}\bar{\pi}_1(t)$. We need the following results:

Lemma 2.1 *If* $\pi_2^\infty = \limsup_{t\to\infty} \pi_2(t)$ *then*

$$\pi_2^\infty \leq \frac{\delta_1 \pi_1^\infty}{\mu + \delta_2}.$$

Proof.

From eqn (2.2) we have

$$\begin{aligned} \frac{d}{dt}\left[\pi_2 \exp[(\mu+\delta_2)t]\right] &= \pi_1 \delta_1 \exp[(\mu+\delta_2)t], \\ &\leq (\pi_1^\infty + \epsilon)\delta_1 \exp[(\mu+\delta_2)t], \quad \forall t \geq t_1(\epsilon) \text{ for any } \epsilon > 0. \end{aligned}$$

Integrating over $[t_1(\epsilon), t]$ gives

$$\begin{aligned} \pi_2(t) &\leq \pi_2(t_1(\epsilon)) \exp[-(\mu+\delta_2)(t-t_1(\epsilon))] + \delta_1(\pi_1^\infty + \epsilon)\left[\frac{1-\exp[-(\mu+\delta_2)(t-t_1(\epsilon))]}{\mu+\delta_2}\right], \\ &\leq \epsilon + \frac{\delta_1(\pi_1^\infty+\epsilon)}{\mu+\delta_2}, \qquad \forall t \geq t_2(\epsilon), \quad \text{for some } t_2(\epsilon) > t_1(\epsilon) \text{ sufficiently large.} \end{aligned}$$

Hence

$$\bar{\pi}_2(t) \leq \epsilon + \frac{\delta_1(\pi_1^\infty+\epsilon)}{\mu+\delta_2}, \qquad \forall t \geq t_2(\epsilon).$$

Letting $t \to \infty$ we have

$$\pi_2^\infty \leq \frac{\delta_1 \pi_1^\infty}{\mu+\delta_2} + \epsilon_1, \qquad \text{where } \epsilon_1 = \epsilon\left(\frac{\mu+\delta_1+\delta_2}{\mu+\delta_2}\right).$$

Suppose that $\pi_2^\infty > (\delta_1\pi_1^\infty)/(\mu+\delta_2)$. Since ϵ_1 is an arbitrary positive constant we can choose $\epsilon_1 = \frac{1}{2}[\pi_2^\infty - (\delta_1\pi_1^\infty)/(\mu+\delta_2)]$. This provides a contradiction and completes the proof.•

Corollary 2.1 *If* $\pi_3^\infty = \limsup_{t\to\infty} \pi_3(t)$ *then*

$$\pi_3^\infty \leq \frac{\delta_1\delta_2\pi_1^\infty}{(\mu+\delta_2)(\mu+\delta_3)}.$$

Proof.

Using eqn (2.3) and following the method of Lemma 2.1 we find that

$$\pi_3^\infty \leq \frac{\delta_2 \pi_2^\infty}{\mu+\delta_3}.$$

The result follows directly.•

Corollary 2.2 *If* $\beta^\infty = \limsup_{t\to\infty} \beta(t)$ *then*

$$\beta^\infty \leq \frac{\pi_1^\infty + \pi_2^\infty + \pi_3^\infty}{\hat{\theta}+\hat{\tau}}.$$

Proof.

Using eqn (2.4) we find that

$$\begin{aligned}
\frac{d\beta}{dt} &\leq \lambda\gamma(\pi_1+\pi_2+\pi_3)-\beta(\lambda\gamma\hat{\theta}+\tau), \\
\text{hence } \frac{d}{dt}\left[\beta\exp[(\lambda\gamma\hat{\theta}+\tau)t]\right] &\leq \lambda\gamma(\pi_1+\pi_2+\pi_3)\exp[(\lambda\gamma\hat{\theta}+\tau)t], \\
&\leq \lambda\gamma(\pi_1^{\infty}+\pi_2^{\infty}+\pi_3^{\infty}+\epsilon)\exp[(\lambda\gamma\hat{\theta}+\tau)t], \text{ for } t\geq t_3(\epsilon).
\end{aligned}$$

The result now follows using the method of Lemma 2.1.•

We now use Lemma 2.1, Corollaries 2.1 and 2.2 together with eqn (2.1) to bound π_1^{∞} above. Suppose that $\pi_1^{\infty}>0$. Given $\epsilon>0$,

$$\begin{aligned}
\frac{d\pi_1}{dt} &\leq (1-\pi_1)\lambda\beta\alpha(1-\phi)-(\mu+\delta_1)\pi_1, \\
&\leq (1-\pi_1)\lambda\alpha(1-\phi)\left(\frac{\pi_1^{\infty}+\pi_2^{\infty}+\pi_3^{\infty}+\epsilon}{\hat{\theta}+\hat{\tau}}\right)-(\mu+\delta_1)\pi_1, \\
&\qquad \forall t\geq t_4(\epsilon) \text{ using Corollary 2.2}, \\
&\leq (1-\pi_1)(\mu+\delta_1)(R_0+\epsilon_2)\pi_1^{\infty}-(\mu+\delta_1)\pi_1, \\
&\qquad \text{where} \quad \epsilon_2=\frac{\lambda\alpha(1-\phi)\epsilon}{(\hat{\theta}+\hat{\tau})(\mu+\delta_1)\pi_1^{\infty}}, \\
&\leq (\mu+\delta_1)[(R_0+\epsilon_2)\pi_1^{\infty}-\pi_1(1+R_0\pi_1^{\infty})].
\end{aligned}$$

Hence

$$\frac{d}{dt}\left[\pi_1(t)\exp[(\mu+\delta_1)(1+R_0\pi_1^{\infty})t]\right] \leq (\mu+\delta_1)(R_0+\epsilon_2)\pi_1^{\infty}\exp[(\mu+\delta_1)(1+R_0\pi_1^{\infty})t]$$
$$\forall t\geq t_4(\epsilon).$$

Following the same method as in Lemma 2.1 we find that

$$\pi_1^{\infty}\leq\epsilon_3+\frac{\pi_1^{\infty}(R_0+\epsilon_2)}{1+R_0\pi_1^{\infty}}=\frac{R_0\pi_1^{\infty}}{1+R_0\pi_1^{\infty}}+\epsilon_4, \tag{2.11}$$

where ϵ_3 is an arbitrarily small positive constant and

$$\epsilon_4=\epsilon_3+\frac{\pi_1^{\infty}\epsilon_2}{1+R_0\pi_1^{\infty}}.$$

When $1\geq R_0\geq 0$ we have that $\pi_1^{\infty}+R_0(\pi_1^{\infty})^2>R_0\pi_1^{\infty}$. This implies that

$$\pi_1^{\infty}-\frac{R_0\pi_1^{\infty}}{1+R_0\pi_1^{\infty}}>0.$$

However since ϵ_3 and ϵ_2 are arbitrary positive constants we can choose

$$\epsilon_4=\frac{1}{2}\left(\pi_1^{\infty}-\frac{R_0\pi_1^{\infty}}{1+R_0\pi_1^{\infty}}\right).$$

This provides a contradiction and hence $\pi_1^\infty = 0$ provided that $1 \geq R_0 \geq 0$.

Let $\liminf_{t\to\infty} \pi_1(t) = \pi_{1,\infty}$ and note that $\pi_1^\infty \geq \pi_{1,\infty} \geq 0$. This implies that $\pi_1^\infty = \pi_{1,\infty} = 0$, and hence $\lim_{t\to\infty} \pi_1(t) = 0$. By Lemma 2.1, Corollary 2.1 and Corollary 2.2 $\pi_1^\infty = 0$ implies that $\pi_2^\infty = \pi_3^\infty = \beta^\infty = 0$ and $\pi_{2,\infty} \geq 0$, $\pi_{3,\infty} \geq 0$ and $\beta_\infty \geq 0$. It follows directly that

$$\lim_{t\to\infty} \pi_2(t) = \lim_{t\to\infty} \pi_3(t) = \lim_{t\to\infty} \beta(t) = 0.$$

This completes the proof of global stability of the disease-free equilibrium when $1 \geq R_0 \geq 0$.•

So far we have examined the behaviour of our model when $R_0 \leq 1$, and as we might have expected we have found that this is a necessary and sufficient condition for the eradication of disease. We now investigate the behaviour of our model when $R_0 > 1$. This is more difficult to deal with analytically and our results are less complete. We first demonstrate that increasing past the threshold of $R_0 = 1$ causes the disease-free equilibrium to become unstable. We then use this result to show that disease will now persist among the population indefinitely. We next show that the endemic equilibrium is locally asymptotically stable when $R_0 > 1$ and unstable when $R_0 < 1$.

Theorem 2.3 *If $R_0 > 1$ then there is still the equilibrium where the disease has died out and this equilibrium is unstable.*

Proof.

Consider the linearised system of eqns (2.1)-(2.4), evaluated at the disease-free equilibrium. This system can be represented in matrix form as

$$\frac{d\mathbf{x}}{dt} = \mathbf{J}\mathbf{x},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta)$ and

$$\mathbf{J} = \begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ \lambda\gamma & \lambda\gamma & \lambda\gamma & -(\lambda\gamma\hat{\theta}+\tau) \end{bmatrix}.$$

We wish to show that at least one eigenvalue of $\mathbf{J}$ has a strictly positive real part. Using the Routh-Hurwitz conditions (May, 1973) it is sufficient to show that the constant

term, a_4, in the characteristic equation of J, $\omega^4 + a_1\omega^3 + a_2\omega^2 + a_3\omega + a_4 = 0$, is strictly negative. It is straightforward to show that

$$\begin{aligned} a_4 &= (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma\hat{\theta}+\tau) \\ &\quad \times \left[1 - \frac{\lambda^2\gamma\alpha(1-\phi)[(\mu+\delta_2)(\mu+\delta_3)+\delta_1(\mu+\delta_3)+\delta_1\delta_2]}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma\hat{\theta}+\tau)}\right], \\ &= (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma\hat{\theta}+\tau)(1-R_0). \end{aligned}$$

Hence if $R_0 > 1$ then $a_4 < 0$ and the result follows.•

Theorem 2.4 *If $R_0 > 1$ and either $\pi(0) > 0$ or $\beta(0) > 0$ then there exists a fixed $\epsilon > 0$ depending only on the model parameters and not the initial conditions such that for some $\eta > 0$*

$$\pi_1 \geq \epsilon\pi_1^\star,\ \pi_2 \geq \epsilon\pi_2^\star,\ \pi_3 \geq \epsilon\pi_3^\star \text{ and } \beta \geq \epsilon\beta^\star, \quad \forall t \geq \eta. \tag{2.12}$$

The proof of this result requires a number of steps. The intuitive argument is that π_1 is the dominant component of $(\pi_1, \pi_2, \pi_3, \beta)$ in the sense that if π_1 becomes small then this causes all of the other components to become small. We shall show that this is indeed the case. From Theorem 2.3 we know that the disease-free equilibrium is unstable when $R_0 > 1$, we shall further show that π_1 cannot become arbitrarily close to zero, and from this we shall deduce that no component can become arbitrarily small. Let $\bar{\pi}_1(t) = \inf_{\xi \geq t} \pi_1(\xi)$, this is monotone increasing in t. Hence given $\epsilon > 0$ there exists $t_5(\epsilon)$ such that $\pi_1(t) \geq \pi_{1,\infty} - \epsilon$ for all $t \geq t_5(\epsilon)$ where $\pi_{1,\infty} = \liminf_{t\to\infty} \pi_1(t)$.

Lemma 2.2 *If $\pi_{2,\infty} = \liminf_{t\to\infty} \pi_2(t)$ then*

$$\pi_{2,\infty} \geq \frac{\delta_1\pi_{1,\infty}}{\mu+\delta_2}.$$

Proof.

From eqn (2.2) we have that

$$\begin{aligned} \frac{d}{dt}\Big[\pi_2 \exp[(\mu+\delta_2)t]\Big] &= \pi_1\delta_1 \exp[(\mu+\delta_2)t], \\ &\geq (\pi_{1,\infty} - \epsilon)\delta_1 \exp[(\mu+\delta_2)t] \quad \forall t \geq t_5(\epsilon), \quad \text{for any } \epsilon > 0. \end{aligned}$$

Integrating over $[t_5(\epsilon), t]$ gives

$$\pi_2(t) \geq \pi_2(t_5(\epsilon)) \exp\Big[-(\mu+\delta_2)(t - t_5(\epsilon))\Big]$$

$$+\delta_1(\pi_{1,\infty}-\epsilon)\left[\frac{1-\exp\left[-(\mu+\delta_2)(t-t_5(\epsilon))\right]}{\mu+\delta_2}\right],$$

$$\geq \frac{\delta_1(\pi_{1,\infty}-\epsilon)}{\mu+\delta_2}-\epsilon \quad \forall t \geq t_6(\epsilon) \text{ for some } t_6(\epsilon) > t_5(\epsilon),$$

sufficiently large.

Hence for all $t \geq t_6(\epsilon)$

$$\bar{\pi}_2(t) \geq \frac{\delta_1(\pi_{1,\infty}-\epsilon)}{\mu+\delta_2}-\epsilon.$$

Letting $t \to \infty$ we get

$$\pi_{2,\infty} \geq \frac{\delta_1\pi_{1,\infty}}{\mu+\delta_2}-\epsilon_1, \qquad \text{where } \epsilon_1 = \frac{\epsilon(\delta_1+\delta_2+\mu)}{\mu+\delta_2}.$$

Suppose that $\pi_{2,\infty} < (\delta_1\pi_{1,\infty})/(\mu+\delta_2)$. Since ϵ_1 is an arbitrary positive constant we can choose $\epsilon_1 = (1/2)[((\delta_1\pi_{1,\infty})/(\mu+\delta_2)) - \pi_{2,\infty}]$. This provides a contradiction and completes the proof of Lemma 2.2.•

Corollary 2.3 *If* $\pi_{3,\infty} = \liminf_{t\to\infty}\pi_3(t)$ *then*

$$\pi_{3,\infty} \geq \frac{\delta_1\delta_2\pi_{1,\infty}}{(\mu+\delta_2)(\mu+\delta_3)}.$$

Proof.

Using eqn (2.3) and following the method of Lemma 2.2 we find that

$$\pi_{3,\infty} \geq \frac{\delta_2\pi_{2,\infty}}{\mu+\delta_3},$$

and the result follows using Lemma 2.2.•

Corollary 2.4 *If* $\beta_\infty = \liminf_{t\to\infty}\beta(t)$ *then*

$$\beta_\infty \geq \frac{\pi_{1,\infty}}{1+\hat{\theta}+\hat{\tau}}.$$

Proof.

Using eqn (2.4) we find that

$$\frac{d\beta}{dt} = \lambda\gamma\pi - \beta\lambda\gamma\pi - \beta\lambda\gamma\hat{\theta} + \beta\lambda\gamma\hat{\theta}\pi - \beta\tau,$$

$$\geq \lambda\gamma\pi_1 - \beta[\lambda\gamma(1+\hat{\theta})+\tau].$$

Hence we have that

$$\frac{d}{dt}\left[\beta\exp\left([\lambda\gamma(1+\hat{\theta})+\tau]t\right)\right] \geq \lambda\gamma\pi_1\exp\left([\lambda\gamma(1+\hat{\theta})+\tau]t\right),$$

and the result follows similarly to Lemma 2.2.•

Lemma 2.3 *Provided that at least one of* $\pi_1(t)$, $\pi_2(t)$, $\pi_3(t)$ *and* $\beta(t)$ *is strictly positive at* $t = 0$ *then* $\pi_1(\Delta t) > 0$, $\pi_2(\Delta t) > 0$, $\pi_3(\Delta t) > 0$ *and* $\beta(\Delta t) > 0$ *for* Δt *small and strictly positive.*

Proof.

We need to consider four separate initial conditions:

1. Suppose that $\beta(0) = 0$. Hence $\pi(0) > 0$. Using a Taylor expansion about $t = 0$ and eqns (2.1)-(2.4) we find that

$$\begin{aligned} \beta(\Delta t) &= \pi(0)\lambda\gamma\Delta t + o(\Delta t) > 0, \\ \text{and} \qquad \pi(\Delta t) &= \pi(0) - (\mu\pi(0) + \delta_3\pi_3(0))\Delta t + o(\Delta t) > 0 \quad \text{(for small } \Delta t\text{)}. \end{aligned}$$

Let $\psi = 1 - \pi$, hence

$$\frac{d\psi}{dt} = -\psi\lambda\beta\alpha(1-\phi) + \mu(1-\psi) + \pi_3\delta_3.$$

If $\pi(0) < 1$ we must have $\psi(0) > 0$, hence $\psi(\Delta t) > 0$ for small enough $\Delta t > 0$, if $\pi(0) = 1$ then $\psi(0) = 0$ and

$$\psi(\Delta t) \geq \mu\pi(0)\Delta t + o(\Delta t) > 0.$$

Hence by choosing $\Delta t > 0$ small enough and starting at $t = \Delta t$ instead of $t = 0$ (if necessary) we can assume that $\pi(0) > 0$, $\psi(0) > 0$ and $\beta(0) > 0$. If $\pi_1(0) = 0$ then $\pi_1(\Delta t) = \psi(0)\lambda\beta(0)\alpha(1-\phi)\Delta t + o(\Delta t) > 0$, if $\Delta t > 0$ is small enough. Hence again by starting at $t = \Delta t$ if necessary we can also assume that $\pi_1(0) > 0$. If $\pi_2(0) = 0$ then $\pi_2(\Delta t) = \delta_1\pi_1(0)\Delta t + o(\Delta t) > 0$, if $\Delta t > 0$ is small enough. Hence we can also assume that $\pi_2(0) > 0$. Similarly if $\pi_3(0) = 0$ then $\pi_3(\Delta t) = \delta_2\pi_2(0)\Delta t + o(\Delta t) > 0$, hence we can also assume that $\pi_3(0) > 0$.

2. Suppose that $\pi(0) = 0$. Hence $\beta(0) > 0$. Following the same method as in the previous case we find that

$$\begin{aligned} \pi(\Delta t) &= \lambda\beta(0)\alpha(1-\phi)\Delta t + o(\Delta t) > 0, && \text{(for small } \Delta t\text{)}, \\ \beta(\Delta t) &= \beta(0) - (\beta(0)\lambda\gamma\hat{\theta} + \beta(0)\tau)\Delta t + o(\Delta t) > 0, && \text{(for small } \Delta t\text{)}, \\ \text{and} \quad \psi(\Delta t) &= 1 - \lambda\beta(0)\alpha(1-\phi)\Delta t + o(\Delta t) > 0, && \text{(for small } \Delta t\text{)}. \end{aligned}$$

Hence by choosing Δt small enough and starting at $t = \Delta t$ we can assume that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$ and as in the previous case we can also assume without loss of generality that $\pi_1(0) > 0$, $\pi_2(0) > 0$ and $\pi_3(0) > 0$.

3. Suppose that $\pi(0) > 0$, $\beta(0) > 0$, and $\psi(0) > 0$. This case is trivial and follows directly as in Case 1.

4. Suppose that $\pi(0) > 0$, $\beta(0) > 0$, and $\psi(0) = 0$. This implies that $\pi(0) = 1$, and hence

$$\psi(\Delta t) \geq \mu \Delta t + o(\Delta t) > 0, \qquad \text{for } \Delta t \text{ small and strictly positive.}$$

Thus it follows directly that by starting at time $t = \Delta t$ we can assume that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$ and the result follows by Case 1.

This completes the proof of Lemma 2.3.•

From Lemma 2.3 we have that there exists fixed ϵ where $1 > \epsilon > 0$ such that if Δt is small enough $\pi_i(\Delta t) \geq \epsilon \pi_i^\star$ for $i = 1, 2, 3$ and $\beta(\Delta t) \geq \epsilon \beta^\star$. Now either $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$ or else $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$. Suppose first that $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$. Then there exists T_1 such that for $t \geq T_1$, $\pi_1 \geq \frac{1}{4}\epsilon\pi_1^\star$, by the definition of $\pi_{1,\infty}$. Then by Lemma 2.2

$$\pi_{2,\infty} \geq \frac{\delta_1 \pi_{1,\infty}}{\mu + \delta_2} \geq \frac{1}{2}\frac{\epsilon\delta_1}{\mu + \delta_2}\pi_1^\star = \frac{1}{2}\epsilon\pi_2^\star.$$

So arguing similarly to above there exists T_2 such that for $t \geq T_2$, $\pi_2 \geq \frac{1}{4}\epsilon\pi_2^\star$. Similarly using Corollary 2.3 we have that there exists T_3 such that for $t \geq T_3$, $\pi_3 \geq \frac{1}{4}\epsilon\pi_3^\star$ and using Corollary 2.4

$$\beta_\infty \geq \frac{\pi_{1,\infty}}{1 + \hat{\theta} + \hat{\tau}} \geq \frac{\frac{1}{2}\epsilon\pi_1^\star}{1 + \hat{\theta} + \hat{\tau}} = \frac{1}{2}\epsilon_1\beta^\star, \qquad \text{say,}$$

where $$\epsilon_1 = \frac{\epsilon}{1 + \hat{\theta} + \hat{\tau}}\frac{\pi_1^\star}{\beta^\star}.$$

Hence there exists T_4 such that for $t \geq T_4$, $\beta \geq \frac{1}{4}\epsilon_1\beta^\star$. Hence if $T = \max\{T_1, T_2, T_3, T_4\}$ and $\tilde{\epsilon} = \min\{\frac{1}{4}\epsilon, \frac{1}{4}\epsilon_1\}$ the results of (2.12) hold with ϵ replaced by $\tilde{\epsilon}$.

Now suppose that $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$, in which case there exists $\zeta \geq \Delta t$ where $\pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star$. Let $t_0 = \inf\{\zeta \geq \Delta t,\ \pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star\}$, and $t_1 = \inf\{\zeta \geq t_0,\ \pi_1(\zeta) > \frac{1}{2}\epsilon\pi_1^\star\}$, where ϵ is fixed and positive. By the definition of t_0 we have that $\pi_1(t_0 + \nu) < \frac{1}{2}\epsilon\pi_1^\star$ if ν is small and positive, hence $t_1 > t_0$. By continuity $\pi_1(t_0) = \pi_1(t_1) = \frac{1}{2}\epsilon\pi_1^\star$, and therefore π_1 is less than $\frac{1}{2}\epsilon\pi_1^\star$ in (t_0, t_1) and greater than $\frac{1}{2}\epsilon\pi_1^\star$ just after t_1. We now show that if π_1 becomes small then π_2, π_3 and β must become small also.

Lemma 2.4 *There exists a time $\bar{T}_1 > 0$ such that if $t_0 + \bar{T}_1 < t_1$ then for all $t \in [t_0 + \bar{T}_1, t_1]$, $0 < \pi_2 < (\frac{1}{2} + \Delta)\pi_2^\star\epsilon$, where Δ and ϵ are small and $\bar{T}_1$ depends only on the model parameters, Δ and ϵ.*

Proof.

In $[t_0, t_1]$ we know that $\pi_1 \leq (1/2)\epsilon\pi_1^*$, hence using eqn (2.2) we have that

$$\frac{d\pi_2}{dt} \leq (1/2)\epsilon\pi_1^*\delta_1 - (\mu + \delta_2)\pi_2,$$

and
$$\frac{d}{dt}\Big[\pi_2 \exp[(\mu+\delta_2)t]\Big] \leq (1/2)\epsilon\pi_1^*\delta_1 \exp[(\mu+\delta_2)t].$$

Integrating over $[t_0, t]$ gives

$$\begin{aligned}
\pi_2 \exp[(\mu+\delta_2)t] - \pi_2(t_0)\exp[(\mu+\delta_2)t_0] &\leq \frac{1}{2}\frac{\epsilon\pi_1^*\delta_1}{\mu+\delta_2}\Big(\exp[(\mu+\delta_2)t] - \exp[(\mu+\delta_2)t_0]\Big), \\
&= \frac{1}{2}\epsilon\pi_2^*\Big(\exp[(\mu+\delta_2)t] - \exp[(\mu+\delta_2)t_0]\Big), \\
\pi_2 &\leq \pi_2(t_0)\exp[-(\mu+\delta_2)(t-t_0)] \\
&\quad + \frac{1}{2}\epsilon\pi_2^*\Big(1 - \exp[-(\mu+\delta_2)(t-t_0)]\Big), \\
\pi_2 &\leq \exp[-(\mu+\delta_2)(t-t_0)] + \frac{1}{2}\epsilon\pi_2^*.
\end{aligned}$$

Hence if Δ is small and positive and t is sufficiently large, say $t \geq t_0 + \bar{T}_1$ where $t_0 + \bar{T}_1 \leq t_1$ then the result follows.•

We have shown that if π_1 is small then this causes π_2 to also become small, we now show similar results for π_3 and β.

Corollary 2.5 *There exists a time $\bar{T}_2 > 0$ such that if $t_0 + \bar{T}_1 + \bar{T}_2 < t_1$ then for all $t \in [t_0 + \bar{T}_1 + \bar{T}_2, t_1]$, $0 < \pi_3 < (\frac{1}{2} + 2\Delta)\pi_3^*\epsilon$, where Δ and ϵ are small and $\bar{T}_2$ depends only on the model parameters, Δ and ϵ.*

Proof.

Similar method to Lemma 2.4 starting with

$$\frac{d\pi_3}{dt} \leq \left(\frac{1}{2} + \Delta\right)\epsilon\pi_2^*\delta_2 - (\mu + \delta_3)\pi_3,$$

and integrating over $[t_0 + \bar{T}_1, t]$.•

Corollary 2.6 *There exists a time $\bar{T}_3 > 0$ such that if $t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3 < t_1$, then for all $t \in [t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3, t_1]$, $0 < \beta < (\frac{1}{2} + 3\Delta)\beta^*\epsilon_1$, where $\epsilon_1 = (\lambda\gamma\pi^* + \lambda\gamma(1-\pi^*)\hat{\theta} + \tau)\epsilon/(\lambda\gamma\hat{\theta} + \tau)$, Δ and ϵ are small and $\bar{T}_3$ depends only on the model parameters, Δ and ϵ.*

Proof.

Similar method to Lemma 2.4 starting with

$$\begin{aligned}\frac{d\beta}{dt} &\leq \lambda\gamma\pi - \beta(\lambda\gamma\hat{\theta} + \tau), \\ &\leq \lambda\gamma\left(\frac{1}{2} + 2\Delta\right)\epsilon\pi^{\star} - \beta(\lambda\gamma\hat{\theta} + \tau), \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_2,\end{aligned}$$

and integrating over $[t_0 + \bar{T}_1 + \bar{T}_2, t]$. For $t \geq t_0 + \bar{T}_1 + \bar{T}_2$ sufficiently large, say $t \geq t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3$ we find that

$$\beta(t) \leq \frac{\lambda\gamma(\frac{1}{2} + 3\Delta)\epsilon\pi^{\star}}{\lambda\gamma\hat{\theta} + \tau}.$$

However it is not necessarily true that $\beta(t) \leq (\frac{1}{2} + 3\Delta)\epsilon\beta^{\star}$ since

$$\beta^{\star} = \frac{\lambda\gamma\pi^{\star}}{\lambda\gamma\pi^{\star} + \lambda\gamma(1 - \pi^{\star})\hat{\theta} + \tau} \leq \frac{\lambda\gamma\pi^{\star}}{\lambda\gamma\hat{\theta} + \tau}.$$

If we now define

$$\epsilon_1 = \frac{[\lambda\gamma\pi^{\star} + \lambda\gamma(1 - \pi^{\star})\hat{\theta} + \tau]\epsilon}{\lambda\gamma\hat{\theta} + \tau},$$

(note that $\epsilon_1 > \epsilon$) then we can again obtain a bound of the required type as our argument shows that $\beta(t) \leq ((1/2) + 3\Delta)\epsilon_1\beta^{\star}$ for $t \geq t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3$.•

We have shown that if π_1 approaches zero then all components must also approach zero. We now show that π_1 cannot become arbitrarily small. We do this by showing that t_1 can be bounded above by a fixed finite value dependent only on the model parameters, ϵ and Δ. Hence π_1 is not below $\frac{1}{2}\epsilon\pi_1^{\star}$ long enough to become arbitrarily close to zero. Now either π_1 is below $\frac{1}{2}\epsilon\pi_1^{\star}$ long enough for all components to become small or π_1 increases past $\frac{1}{2}\epsilon\pi_1^{\star}$ before all components become small. Hence we have that either (i) $t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3]$, or (ii) $t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3]$. We wish to show that $t_1 < T$ where T is a fixed finite value dependent only on the model parameters, ϵ and Δ. If case (ii) is true then we are finished. Case (i) is where all components become small before time t_1 and it is this situation we now deal with. Since the disease-free equilibrium is unstable we can use this to show that π_1 cannot stay small indefinitely. The following two results deal with this issue:

Corollary 2.7 *Let $F_1(\omega, \epsilon)$ be an n^{th} degree polynomial in ω and ϵ. Denote the (possibly complex) roots of $F_1(\omega, \epsilon) = 0$ by $\omega_j(\epsilon)$ for $j = 1, \ldots, n$. Then each $\omega_j(\epsilon)$ is defined and continuous in ϵ in a neighbourhood of $\epsilon = 0$.*

Proof.

$F_1(\omega, \epsilon)$ is a polynomial and therefore it is analytic in a neighbourhood of (0,0), the result follows directly from Corollary 6.6 in Chow and Hale (1982).•

Our next Lemma is a key point in the argument. As previously stated we use the instability of the disease-free equilibrium to show that π_1 cannot stay small indefinitely and in fact rises again to $\frac{1}{2}\epsilon\pi_1^\star$ by a time which is finite and depends only on Δ, ϵ and the model parameters.

Lemma 2.5 *If $\pi_1(t)$ drops to below $\frac{1}{2}\epsilon\pi_1^\star$ at time t_0 then $\pi_1(t)$ returns to $\frac{1}{2}\epsilon\pi_1^\star$ by at least time $t_1^+ = t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3, t_2 + \bar{T}_4]$ where $t_1^+ - t_0$ is finite and depends only on Δ, ϵ and the model parameters.*

Proof.

Suppose that ϵ_2 is real and positive and $1 \geq \epsilon_2 \geq 0$ and consider the matrix

$$\mathbf{J}(\epsilon_2) = \begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda\alpha(1-\phi)(1-\epsilon_2) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ \lambda\gamma & \lambda\gamma & \lambda\gamma & -(\lambda\gamma(\epsilon_2+\hat{\theta})+\tau) \end{bmatrix}.$$

When $\epsilon_2 = 0$, $\mathbf{J}(0) = \mathbf{J}$, the linearised stability matrix about the disease-free equilibrium as used in the proof of Theorem 2.3. Denote the eigenvalues of $\mathbf{J}(\epsilon_2)$ by $w_1(\epsilon_2)$, $w_2(\epsilon_2)$, $w_3(\epsilon_2)$ and $w_4(\epsilon_2)$. For M large and positive we have that $\mathbf{J}(\epsilon_2) + M\mathbf{I}$ is a non-negative irreducible matrix. Using Lemma 2.1 from Nold (1980), (the Perron-Froebenius Theorem), the characteristic equation of this matrix has a simple root equal to its spectral radius. The eigenvalues of $\mathbf{J}(\epsilon_2)+M\mathbf{I}$ are $M+w_1(\epsilon_2)$, $M+w_2(\epsilon_2)$, $M+w_3(\epsilon_2)$ and $M+w_4(\epsilon_2)$. Hence if $M+w_1(\epsilon_2)$ is the spectral radius of $\mathbf{J}(\epsilon_2)+M\mathbf{I}$ then $M+w_1(\epsilon_2)$ is real and all other eigenvalues of $\mathbf{J}(\epsilon_2)+M\mathbf{I}$ have strictly smaller real parts. Hence $w_1(\epsilon_2)$ is real and the other eigenvalues of $\mathbf{J}(\epsilon_2)$ have strictly smaller real parts. In particular this is true for $\epsilon_2 = 0$. Moreover from Corollary 2.7 we have that the roots of the characteristic equation of $\mathbf{J}(\epsilon_2)$ are continuous functions of ϵ_2, hence $w_1(\epsilon_2) \to w_1(0)$ as $\epsilon_2 \to 0$. From the proof of Theorem 2.3 we know that $w_1(0) > 0$ if $R_0 > 1$. Therefore by choosing ϵ_2 small enough we can ensure that $w_1(\epsilon_2) > 0$. Without loss of generality we can assume that $1 > \epsilon_2 > 0$. We can choose ϵ small enough such that

$$\frac{1}{2}\epsilon\pi_1^\star + \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star + \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star < \epsilon_2.$$

Hence using Lemma 2.4 and Corollary 2.5 for $t_1 > t \geq t_0 + \bar{T}_1 + \bar{T}_2$ we have that $\pi_1 + \pi_2 + \pi_3 < \epsilon_2$. Let $t_2 = \inf\{\zeta : \text{ for } t_1 > t \geq t_0 + \zeta,\ \pi(t) < \epsilon_2\}$, and hence either $t_2 = 0$ or $\pi(t_0 + t_2) = \epsilon_2$ and $t_0 + t_2$ is the last time before t_1 that $\pi(t) \geq \epsilon_2$ and note that $t_2 \leq \bar{T}_1 + \bar{T}_2$. If $t_1 < t_0 + \bar{T}_1 + \bar{T}_2$ then we have the desired result. Now we consider the case where $t_1 \geq t \geq t_0 + \bar{T}_1 + \bar{T}_2$. From equations (2.1)-(2.4) we have that for $t_1 \geq t \geq t_0 + \bar{T}_1 + \bar{T}_2$

$$\begin{aligned}
\frac{d\pi_1}{dt} &\geq (1-\epsilon_2)\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\pi_1, \\
\frac{d\pi_2}{dt} &= \delta_1\pi_1 - (\mu+\delta_2)\pi_2, \\
\frac{d\pi_3}{dt} &= \delta_2\pi_2 - (\mu+\delta_3)\pi_3, \\
\text{and} \qquad \frac{d\beta}{dt} &\geq \lambda\gamma(\pi_1+\pi_2+\pi_3) - (\lambda\gamma(\epsilon_2+\hat{\theta})+\tau)\beta.
\end{aligned}$$

Hence

$$\frac{d\mathbf{x}}{dt} \geq \mathbf{J}(\epsilon_2)\mathbf{x},$$

where $\mathbf{x} = (\pi_1, \pi_2, \pi_3, \beta)^T$. From Lemma 2.1 in Nold (1980), $\mathbf{J}(\epsilon_2)$ has a strictly positive left eigenvector, $\mathbf{e} = (e_1, e_2, e_3, e_4)$ corresponding to its spectral radius $w_1(\epsilon_2)$. Hence

$$\mathbf{e}\,\frac{d\mathbf{x}}{dt} \geq \mathbf{e}\,\mathbf{J}(\epsilon_2)\mathbf{x} = w_1(\epsilon_2)\,\mathbf{e}\,\mathbf{x}.$$

Thus integrating over $[t_0 + t_2, t]$ gives

$$\begin{aligned}
\mathbf{e}.\,\mathbf{x}(t) &\geq \mathbf{e}.\,\mathbf{x}(t_0+t_2)\exp[w_1(\epsilon_2)(t-t_0-t_2)], \\
&\geq \left(e_1\pi_1(t_0+t_2) + e_2\pi_2(t_0+t_2) + e_3\pi_3(t_0+t_2)\right)\exp[w_1(\epsilon_2)(t-t_0-t_2)], \\
&\geq \pi(t_0+t_2)\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)], \\
&\begin{cases} = \epsilon_2\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)], & \text{if } t_2 > 0, \\ \geq \frac{1}{2}\epsilon\pi_1^\star\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)], & \text{if } t_2 = 0. \end{cases}
\end{aligned}$$

Therefore after a time $t_0 + t_2 + \bar{T}_4$ and provided that $t_1 \geq t_0 + t_2 + \bar{T}_4$ we have that

$$\mathbf{e}.\,\mathbf{x}(t) > \mathbf{e}.\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_1\beta^\star\right), \tag{2.13}$$

where $\bar{T}_4$ depends only on $\epsilon_1, \epsilon, \Delta$ and the model parameters. We have also shown that provided that $t_0 \leq t \leq t_1$ then after a time $t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3]$ we have that $\pi_1(t) \leq \frac{1}{2}\epsilon\pi_1^\star$, $\pi_2(t) \leq (\frac{1}{2}+\Delta)\epsilon\pi_2^\star$, $\pi_3(t) \leq (\frac{1}{2}+2\Delta)\epsilon\pi_3^\star$ and $\beta(t) \leq (\frac{1}{2}+3\Delta)\epsilon_1\beta^\star$. Hence

$$\mathbf{e}.\,\mathbf{x}(t) \leq \mathbf{e}.\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_1\beta^\star\right).$$

Therefore if $t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3, t_2 + \bar{T}_4]$ we have a contradiction to eqn (2.13) and hence $t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3, t_2 + \bar{T}_4]$. This concludes the proof of Lemma 2.5.•

We have shown that the first time π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$ it must return back to this level after a duration of at most $T = \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_1 + \bar{T}_2 + \bar{T}_3, t_2 + \bar{T}_4]$ later. By shifting the time origin it is easy to see that the argument can be extended to cover *any* time π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$. Hence if π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$ at $\tilde{t}_0$ then for $t \in [\tilde{t}_0, \tilde{t}_0 + T]$ we have that

$$\begin{aligned} \frac{d\pi_1}{dt} &\geq -(\mu + \delta_1)\pi_1. \\ \text{Integrating we deduce that} \qquad \pi_1 &\geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu+\delta_1)(t-\tilde{t}_0)], \\ &\geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu+\delta_1)T], \end{aligned}$$

where T is a fixed duration dependent only on $\epsilon, \epsilon_2, \Delta$ and the model parameters. Since $(1/2)\epsilon\pi_1^\star \exp[-(\mu+\delta_1)T]$ is strictly positive we have that $\pi_{1,\infty} > 0$. Hence by the remark on p.58 (2.12) is true and therefore by reducing ϵ we have that there exists a fixed lower bound $\epsilon > 0$ and $\eta > 0$ such that for all $t \geq \eta$, $\pi_1(t) \geq \epsilon$, $\pi_2(t) \geq \epsilon$, $\pi_3(t) \geq \epsilon$ and $\beta(t) \geq \epsilon$. This completes the proof of Theorem 2.4.•

Theorem 2.5 *The endemic equilibrium* $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star)$ *is locally stable if* $R_0 > 1$.

Proof.

The Jacobian matrix for our model evaluated at the endemic equilibrium $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star)$, is

$$\mathbf{J} = \begin{bmatrix} -(\mu+\delta_1)-\lambda\beta^*\alpha(1-\phi) & -\lambda\beta^*\alpha(1-\phi) & -\lambda\beta^*\alpha(1-\phi) & (1-\pi^*)\lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ (1-\beta^*(1-\hat{\theta}))\lambda\gamma & (1-\beta^*(1-\hat{\theta}))\lambda\gamma & (1-\beta^*(1-\hat{\theta}))\lambda\gamma & -(\hat{\theta}+(1-\hat{\theta})\pi^*+\hat{\tau})\lambda\gamma \end{bmatrix}.$$

If the characteristic equation of $\mathbf{J}$ is denoted by $\omega^4 + a_1\omega^3 + a_2\omega^2 + a_3\omega + a_4 = 0$ then using the Routh-Hurwitz conditions for a quartic polynomial we require to show that $a_i > 0$ for $i = 1, 2, 3, 4$, and $a_1a_2a_3 > a_3^2 + a_1^2a_4$. This is straightforward but requires a considerable amount of algebra, see Appendix A for details.•

2.5.1 Comments on Global Stability

We have been unable to show that when $R_0 > 1$ and disease is initially present our model tends to the endemic equilibrium. However we can use our previous persistence result to derive additional sufficient conditions for global stability. The current coordinate system of our model is $(\pi_1, \pi_2, \pi_3, \beta)$. We do not consider this system directly. Instead we use a translated form of the original coordinate system where the origin in this new system corresponds to $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star)$ in the original form. This translation gives us a new set of model equations,

$$\frac{d\tilde{\pi}_1}{dt} = (1-\pi^\star)\lambda\tilde{\beta}\alpha(1-\phi) - \tilde{\pi}\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\tilde{\pi}_1, \tag{2.14}$$

$$\frac{d\tilde{\pi}_2}{dt} = \delta_1\tilde{\pi}_1 - (\mu+\delta_2)\tilde{\pi}_2, \tag{2.15}$$

$$\frac{d\tilde{\pi}_3}{dt} = \delta_2\tilde{\pi}_2 - (\mu+\delta_3)\tilde{\pi}_3, \tag{2.16}$$

and

$$\frac{d\tilde{\beta}}{dt} = [1-\beta^\star+\beta^\star\hat{\theta}]\lambda\gamma\tilde{\pi} - \tilde{\beta}\lambda\gamma[\hat{\theta}+(1-\hat{\theta})\pi] - \tilde{\beta}\tau, \tag{2.17}$$

where $\tilde{\beta} = \beta - \beta^\star$, $\tilde{\pi}_i = \pi_i - \pi_i^\star$ for $i = 1,2,3$, and $\tilde{\pi} = \tilde{\pi}_1 + \tilde{\pi}_2 + \tilde{\pi}_3$. In this new system we wish to show that $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}) \to (0,0,0,0)$ as $t \to \infty$, which is equivalent to $(\pi_1, \pi_2, \pi_3, \beta) \to (\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star)$ as $t \to \infty$. We can write the system represented by eqns (2.14)-(2.17) in matrix form as $d\tilde{\mathbf{x}}/dt = \mathbf{V}(\mathbf{x})\tilde{\mathbf{x}}$, where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta)$, $\tilde{\mathbf{x}}^T = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta})$ and

$$\mathbf{V}(\mathbf{x}) = \begin{bmatrix} -(\mu+\delta_1)-\lambda\beta\alpha(1-\phi) & -\lambda\beta\alpha(1-\phi) & -\lambda\beta\alpha(1-\phi) & (1-\pi^\star)\lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ (1-\beta^\star+\beta^\star\hat{\theta})\lambda\gamma & (1-\beta^\star+\beta^\star\hat{\theta})\lambda\gamma & (1-\beta^\star+\beta^\star\hat{\theta})\lambda\gamma & -(\hat{\theta}+(1-\hat{\theta})\pi+\hat{\tau})\lambda\gamma \end{bmatrix}.$$

When $\mathbf{x} = \mathbf{0}$ we have that

$$\mathbf{V}(\mathbf{0}) = \begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & (1-\pi^\star)\lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ (1-\beta^\star+\beta^\star\hat{\theta})\lambda\gamma & (1-\beta^\star+\beta^\star\hat{\theta})\lambda\gamma & (1-\beta^\star+\beta^\star\hat{\theta})\lambda\gamma & -(\hat{\theta}+\hat{\tau})\lambda\gamma \end{bmatrix},$$

where the only strictly negative entries are on the leading diagonal. We are also interested in an additional coordinate system, that of $(\tilde{\pi}, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta})$. This system is easily obtained from the $(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta})$ system by adding eqns (2.14)-(2.16) and replacing $\tilde{\pi}_1$

with $\tilde{\pi} - \tilde{\pi}_2 - \tilde{\pi}_3$. Thus we have that $d\tilde{\mathbf{y}}/dt = \mathbf{W}(\mathbf{y})\tilde{\mathbf{y}}$, where $\mathbf{y}^T = (\pi, \pi_2, \pi_3, \beta)$, $\tilde{\mathbf{y}}^T = (\tilde{\pi}, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta})$ and

$$\mathbf{W}(\mathbf{y}) = \begin{bmatrix} -\mu-\lambda\beta\alpha(1-\phi) & 0 & -\delta_3 & (1-\pi^*)\lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_2+\delta_1) & -\delta_1 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ (1-\beta^*+\beta^*\hat{\theta})\lambda\gamma & 0 & 0 & -(\hat{\theta}+(1-\hat{\theta})\pi+\hat{\tau})\lambda\gamma \end{bmatrix}.$$

Notice that in $\mathbf{W}(\mathbf{y})$ the variables $\pi(t)$ and $\beta(t)$ appear only on the leading diagonal. The two matrices $\mathbf{V}(\mathbf{x})$ and $\mathbf{W}(\mathbf{y})$ have different structures each with its own useful property. The $\mathbf{V}(\mathbf{x})$ form has *non-negative* entries except on the leading diagonal when $\mathbf{x} = \mathbf{0}$. The $\mathbf{W}(\mathbf{y})$ form has *constant* entries except on the leading diagonal. These two matrices share an important property shown in the following lemma.

Lemma 2.6 $\mathbf{V}(\mathbf{x})$ *and* $\mathbf{W}(\mathbf{y})$ *have the same eigenvalues.*

Proof.

$\tilde{\mathbf{y}} = \mathbf{J}\tilde{\mathbf{x}}$ where

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It is straightforward to verify that $\mathbf{V}(\mathbf{x}) = \mathbf{J}^{-1}\mathbf{W}(\mathbf{y})\mathbf{J}$. Thus if $\mathbf{e}$ is a right eigenvector of $\mathbf{V}(\mathbf{x})$ with corresponding eigenvalue ω we must have that $\mathbf{Je}$ is a right eigenvector of $\mathbf{W}(\mathbf{y})$ with corresponding eigenvalue ω. Similarly if $\mathbf{f}$ is a right eigenvector of $\mathbf{W}(\mathbf{y})$ with eigenvalue ω, $\mathbf{J}^{-1}\mathbf{f}$ is a right eigenvector of $\mathbf{V}(\mathbf{x})$ with the same eigenvalue. Hence $\mathbf{V}(\mathbf{x})$ and $\mathbf{W}(\mathbf{y})$ have the same eigenvalues.•

Using Theorem 2.4 we can replace the variables $\pi(t)$ and $\beta(t)$ in $\mathbf{W}(\mathbf{y})$ with a constant lower bound, ϵ. Hence if we define

$$\mathbf{W}^+ = \begin{bmatrix} -\mu-\lambda\alpha(1-\phi)\epsilon & 0 & -\delta_3 & (1-\pi^*)\lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_1+\delta_2) & -\delta_1 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ (1-\beta^*+\beta^*\hat{\theta})\lambda\gamma & 0 & 0 & -(\hat{\theta}+(1-\hat{\theta})\epsilon+\hat{\tau})\lambda\gamma \end{bmatrix},$$

then note that for $t \geq \eta$ $\mathbf{W(y)} \leq \mathbf{W^+}$. We can write $\mathbf{W^+}$ as $\mathbf{W(0)} - \epsilon\mathbf{E}$ where

$$\mathbf{E} = \begin{bmatrix} \lambda\alpha(1-\phi) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda\gamma(1-\hat{\theta}) \end{bmatrix}.$$

From Lemma 2.6 the eigenvalues of $\mathbf{W(0)}$ are the same as those of $\mathbf{V(0)}$. Moreover if M is large enough $\mathbf{V(0)}+M\mathbf{I}$, (where $\mathbf{I}$ is the identity matrix) is an irreducible matrix with non-negative elements and has a unique strictly positive eigenvector $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star)$. From Lemma 2.1 in Nold (1980) we find that the eigenvalue corresponding to the eigenvector $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star)$ is a simple eigenvalue and is also the spectral radius of $\mathbf{V(0)} + M\mathbf{I}$. Hence all eigenvalues of $\mathbf{V(0)} + M\mathbf{I}$ lie in a circle centered on the origin with radius M. Thus all eigenvalues of $\mathbf{V(0)}$ and hence $\mathbf{W(0)}$ lie in a circle centered on $(-M, 0)$ with radius M. Moreover zero is a simple eigenvalue of $\mathbf{W(0)}$.

Consider the characteristic equation of $\mathbf{W^+} = \mathbf{W(0)} - \epsilon\mathbf{E}$, this is of the form

$$\omega^4 + a_1(\epsilon)\omega^3 + a_2(\epsilon)\omega^2 + a_3(\epsilon)\omega + a_4(\epsilon) = 0, \tag{2.18}$$

where $a_1(\epsilon)$, $a_2(\epsilon)$, $a_3(\epsilon)$, $a_4(\epsilon)$ are continuous functions of ϵ. After some algebra we find that

$$\begin{aligned} a_4(\epsilon) &= (\mu + \lambda\alpha(1-\phi)\epsilon)(\lambda\gamma(\hat{\theta} + (1-\hat{\theta})\epsilon) + \tau)[(\delta_1 + \delta_2 + \mu)(\delta_3 + \mu) + \delta_1\delta_2] \\ &\quad + \delta_1\delta_2\delta_3(\lambda\gamma(\hat{\theta} + (1-\hat{\theta})\epsilon) + \tau) \\ &\quad - (1-\pi^\star)\lambda\alpha\lambda\gamma(1-\phi)[1 - \beta^\star + \beta^\star\hat{\theta}][(\delta_1 + \delta_2 + \mu)(\delta_3 + \mu) + \delta_1\delta_2]. \end{aligned} \tag{2.19}$$

When $\epsilon = 0$ we have that $\mathbf{W^+} = \mathbf{W(0)}$ and we know that zero is an eigenvalue of $\mathbf{W(0)}$. Let the roots of eqn (2.18) be denoted by $\{\omega_1, \omega_2, \omega_3, \omega_4\}$. When $\epsilon = 0$ three eigenvalues ω_1, ω_2 and ω_3 say, have real parts which are strictly negative and the fourth eigenvalue ω_4 is real and lies at the origin. Now consider $\epsilon > 0$. From Corollary 2.7 we know that the eigenvalues are continuous in ϵ in a neighbourhood about the origin. By continuity ω_1, ω_2 and ω_3 must still have strictly negative real parts for ϵ small. Now suppose that ω_4 has a non-negative real part for ϵ small. Hence ω_4 must be real and $a_4 = \omega_1\omega_2\omega_3\omega_4 < 0$ for small ϵ, however this is impossible since $a_4(0) = 0$ and $a_4(\epsilon)$ is increasing in ϵ. Hence for ϵ small and positive all eigenvalues of $\mathbf{W^+}$ have strictly negative real parts. So all eigenvalues of $\mathbf{W^{+T}}$ also have strictly negative real parts.

The well known theorem of Lyapunov (1892) states that all the eigenvalues of a matrix $\mathbf{A}$ have negative real parts if and only if there exists a symmetric positive

definite matrix $\mathbf{P}$ such that $\mathbf{AP}+\mathbf{PA}^T$ is negative definite. Hence we have that there exists a symmetric positive definite matrix $\mathbf{P}$ such that $\mathbf{W}^{+T}\mathbf{P}+\mathbf{PW}^{+}=-\mathbf{Q}$ where $\mathbf{Q}$ is a positive definite matrix. This result is not sufficient for us to show global stability of the endemic equilibrium although we can do so if $\mathbf{P}$ has the form

$$\mathbf{P}=\begin{bmatrix} P_{11} & 0 & 0 & 0 \\ 0 & P_{22} & P_{23} & 0 \\ 0 & P_{32} & P_{33} & 0 \\ 0 & 0 & 0 & P_{44} \end{bmatrix},$$

where P_{11} and P_{44} are strictly positive. If $\mathbf{P}$ is of this form then $v=\tilde{\mathbf{y}}^T\mathbf{P}\tilde{\mathbf{y}}$ is a Lyapunov function for our system since v is always positive and

$$\begin{aligned} \frac{dv}{dt} &= \tilde{\mathbf{y}}^T(\mathbf{W}(\mathbf{y})^{\mathbf{T}}\mathbf{P}+\mathbf{PW}(\mathbf{y}))\tilde{\mathbf{y}}, \\ &\leq \tilde{\mathbf{y}}^T(\mathbf{W}^{+T}\mathbf{P}+\mathbf{PW}^{+})\tilde{\mathbf{y}}, \qquad \text{due to the form of } \mathbf{P}, \\ &= -\tilde{\mathbf{y}}^T\mathbf{Q}\tilde{\mathbf{y}}, \\ &\leq -\omega_{\min}(\mathbf{Q})|\tilde{\mathbf{y}}|^2, \end{aligned}$$

where $\omega_{\min}(\mathbf{Q})$ is the smallest (strictly positive) eigenvalue of $\mathbf{Q}$,

$$\leq -\frac{\omega_{\min}(\mathbf{Q})}{\omega_{\max}(\mathbf{P})}\tilde{\mathbf{y}}^T\mathbf{P}\tilde{\mathbf{y}},$$

where $\omega_{\max}(\mathbf{P})$ is the largest (strictly positive) eigenvalue of $\mathbf{P}$,

$$= -\frac{\omega_{\min}(\mathbf{Q})}{\omega_{\max}(\mathbf{P})}v,$$

and $\omega_{\min}(\mathbf{Q})$ and $\omega_{\max}(\mathbf{P})$ are both strictly positive. Hence

$$0\leq v\leq v(0)\exp\left[-\frac{\omega_{\min}(\mathbf{Q})}{\omega_{\max}(\mathbf{P})}t\right] \qquad \to 0 \quad \text{as} \quad t\to\infty.$$

So as

$$|\tilde{\mathbf{y}}|^2\leq\frac{1}{\omega_{\min}(\mathbf{P})}\tilde{\mathbf{y}}^T\mathbf{P}\tilde{\mathbf{y}}, \qquad \tilde{\mathbf{y}}\to 0 \quad \text{as} \quad t\to\infty,$$

where $\omega_{\min}(\mathbf{P})$ is the smallest (strictly positive) eigenvalue of $\mathbf{P}$. In particular we have global asymptotic stability when for some positive definite matrix $\mathbf{Q}$ we can choose $\mathbf{P}$ a positive diagonal matrix. Therefore we require the (generally stronger) property that $\mathbf{W}^{+}$ is diagonally stable rather than Lyapunov stable.

Barker et al., (1978) discuss various stability type conditions on a matrix $\mathbf{A}$ related to the consistency of the Lyapunov equation $\mathbf{AD}+\mathbf{DA}^{\mathbf{T}}=\mathbf{Q}$ where $\mathbf{D}$ is a positive

diagonal matrix and $\mathbf{Q}$ is positive definite. In particular they show that such a $\mathbf{D}$ exists for any matrix $\mathbf{A} = (a_{ij})$ where $\mathbf{A}$ has a positive diagonal and $\mathcal{M}(\mathbf{A})$ is a nonsingular M-matrix where

$$\mathcal{M}_{ij} = \begin{cases} |a_{ij}| & \text{if } i = j, \\ -|a_{ij}| & \text{if } i \neq j. \end{cases}$$

Recall that a nonsingular M-matrix is a matrix whose off diagonal entries are less than or equal to zero and all its principal minors are strictly positive (Barker et al., 1978).

By considering the matrix $\mathbf{S} = -\mathbf{W}^{+T}$ we find that

$$\mathcal{M}(\mathbf{S}) = \begin{bmatrix} \mu+\lambda\alpha(1-\phi)\epsilon & -\delta_1 & 0 & -(1-\beta^*+\beta^*\hat{\theta})\lambda\gamma \\ 0 & \mu+\delta_1+\delta_2 & -\delta_2 & 0 \\ -\delta_3 & -\delta_1 & \mu+\delta_3 & 0 \\ -(1-\pi^*)\lambda\alpha(1-\phi) & 0 & 0 & (\hat{\theta}+(1-\hat{\theta})\epsilon+\hat{\tau})\lambda\gamma \end{bmatrix}.$$

We now require to show that all the principal minors are strictly positive. After some algebra we find that sufficient conditions for this are that $\det \mathcal{M}(\mathbf{S}) =$

$$\Big\{\Big[(\mu+\lambda\alpha(1-\phi)\epsilon)(\hat{\theta}+(1-\hat{\theta})\epsilon+\hat{\tau}) - (1-\beta^*+\beta^*\hat{\theta})(1-\pi^*)\lambda\alpha(1-\phi)\Big]$$

$$\times \quad \lambda\gamma\Big[(\mu+\delta_1+\delta_2)(\mu+\delta_3) - \delta_1\delta_2\Big]\Big\} - \delta_1\delta_2\delta_3(\hat{\theta}+(1-\hat{\theta})\epsilon+\hat{\tau})\lambda\gamma > 0, \quad (2.20)$$

and $(\mu+\delta_1+\delta_2)(\mu+\delta_3) - \delta_1\delta_2 > 0$. This second condition is satisfied for most parameter values in the literature. It will always be true if $\delta_3 > \delta_2$ or $\delta_3 > \delta_1$. Hence if these conditions are satisfied then there exists a positive diagonal matrix $\mathbf{D}$ such that $\mathbf{W}^{+T}\mathbf{D} + \mathbf{D}\mathbf{W}^{+}$ is negative definite and therefore if disease is initially present the system will tend to the endemic equilibrium. However unlike the sufficient condition of $a_4(\epsilon) > 0$ (shown in eqn (2.19)) for Lyapunov stability of $\mathbf{W}^{+}$, the inequality (2.20) is not satisfied near $\epsilon = 0$ (but could be satisfied for ϵ larger). In particular when $\delta_1 = 0$ $\mathbf{W}^{+}$ is both Lyapunov and diagonally stable. However when $\delta_1 = 0$ our model simplifies down to the original model discussed by Kaplan and O'Keefe (1993). Intuitively we expect that the lower bound on ϵ will be close to zero for R_0 close to one which suggests that our argument, while not strong enough to show that the disease tends to the unique endemic equilibrium if it is initially present for all values of $R_0 > 1$, may hold true for larger values of R_0.

2.6 Simulation Study of the Simple Model

In the previous section we discussed in detail key theoretical properties of the Simple Model. We used mathematical results to analyse the long term behaviour of this model and established the conditions necessary for the disease to die out in the population or become endemic. We now demonstrate the behaviour of the Simple Model graphically using simulation. Simulation allows us to validate our previous theoretical results and puts our model onto a more practical footing by demonstrating how the disease progresses through the population over time.

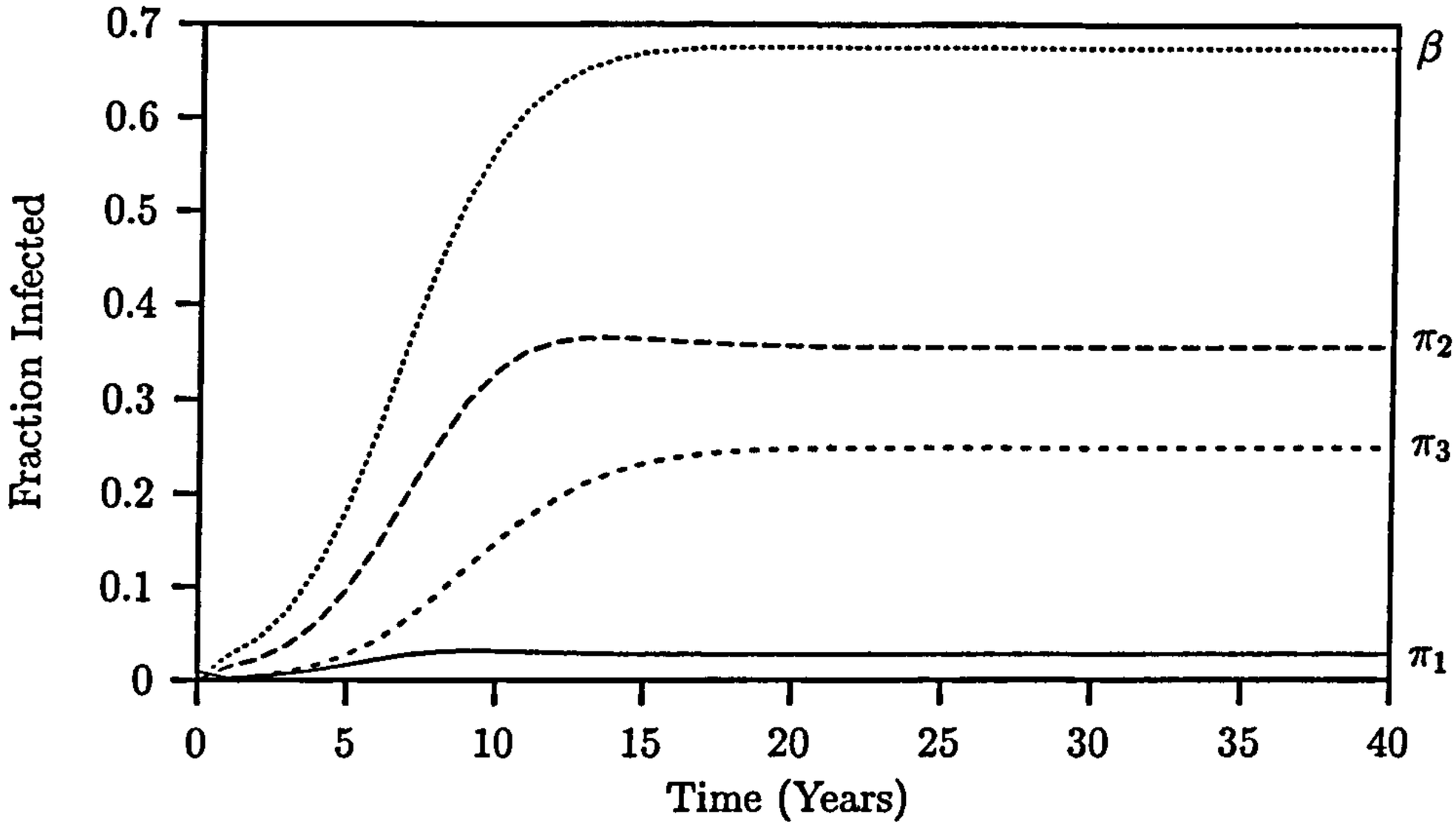
We now use the SOLVER numerical integration package to produce estimates of the prevalence of disease over time given by eqns (2.1)-(2.4). However we first need to estimate the parameters mentioned in Assumptions 1-9. For example we require an estimate of the shared injection rate and the probability with which addicts successfully clean needles prior to use. Table 2.1 contains a summary of the parameter estimates we use to simulate our current model. Justifications for each of these estimates can be found in Appendix B.

Table 2.1: Summary of Parameter Estimates

Parameter	Estimate
λ	246.22 per year
γ	0.90797
α	0.00601 per shared injection
μ	0.1333 per year
ϕ	0.64
θ	0.0
τ	15.531 per year
δ_1	4.6154 per year
δ_2	0.2281 per year
δ_3	0.1920 per year

We are interested in demonstrating two key properties of the Simple Model. Firstly, if the parameter estimates are such that $R_0 > 1$ then provided that disease is initially present in at least one individual addict or needle, then HIV will spread among the population until a steady state is reached where a fraction $\pi^\star$ of all addicts and a

Figure 2.1: System Tends to Endemic Equilibrium when $R_0 > 1$



fraction $\beta^\star$ of all needles are infected with HIV. Secondly, if the parameter estimates are such that $R_0 \leq 1$ then the disease will die out in both addicts and needles and the system will approach the disease-free steady state. To illustrate these two properties we use two sets of parameter estimates, the first set gives rise to a value of $R_0 = 3.596 > 1$ and in the second $R_0 = 0.903 < 1$. The first set of parameter estimates are as shown in Table 2.1 and the second set are the same as the first except that ϕ, the probability that an addict successfully cleans a needle prior to use, is increased from 0.64 to 0.887. This reduces reduces R_0 from 3.596 to 0.903. We now investigate the behaviour of the Simple Model for these two sets of parameters.

Parameter Set One - $R_0 = 3.596$

We now simulate the Simple Model using a set of parameter estimates where $R_0 = 3.596$. Figure 2.1 shows the Simple Model simulated over forty years. At time zero we have assumed that one percent of the total population of addicts are in stage one infectivity, at this time no other addicts or needles are infectious. The lower three lines on the figure show the behaviour over time of the three stages of infectivity among the addict population, the uppermost line represents the total fraction of needles infected with HIV. It is clear that the fraction of addicts infected in each stage eventually reaches a steady state as does the fraction of infected needles. We can also observe that the fraction of addicts in stage one infectivity reaches a steady state first with the

other stages reaching a steady state later. The approximate steady state values are $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star) = (0.027, 0.355, 0.249, 0.675)$, which correspond to $\pi^\star = 0.6326$.

Figures 2.2-2.4 examine the behaviour of the Simple Model using the same set of parameters as in Figure 2.1 but with six sets of different initial conditions. The six starting conditions of the system were $(0.3, 0)$, $(0.7, 0)$, $(1.0, 0.3)$, $(0.7, 1.0)$, $(0.3, 1.0)$ and $(0, 0.3)$, where (x, y) means that $\pi(0) = x$ and $\beta(0) = y$. In the case of the addicts it was assumed that all initially infectious addicts were in stage one infectivity, (i.e. $\pi(0) = 0.3$ means that $\pi_1(0) = 0.3$). Each plot shows the progress of the total fraction of infected addicts and the total fraction of infected needles as time progresses for each of the different initial conditions. It is clear that in each case the level of disease in both addicts and needles reaches a steady state value and moreover this value is the same for each of the six initial conditions. Hence these simulations suggest that for the current set of parameter estimates the total long term prevalence of disease in addicts and needles is 0.6326 and 0.675 respectively. An interesting feature of these simulations is the very similar behaviour of the fraction of infected needles and the fraction of infected addicts. It turns out that whatever the initial conditions the prevalence of disease in needles appears to very quickly reach a quasi-steady state relationship determined by the prevalence of disease among addicts. This is a useful property which we discuss in the following chapter. In summary our simulations (together with others not illustrated) suggest that if disease is initially present then the model will tend to the unique endemic equilibrium.

Parameter Set Two - $R_0 = 0.903$

We now simulate the Simple Model using a set of parameter estimates where $R_0 = 0.903$. Figure 2.5 shows the Simple Model simulated over 120 years. At time zero we have assumed that the population is in an endemic steady state where 2.7% of the total population of addicts are in stage one infectivity, 35.5% are in stage two infectivity and 24.9% are in stage three infectivity. We also assume that 67.5% of needles are infected with HIV. These values correspond to the endemic steady state shown in Figure 2.1. We suppose that at time zero R_0 has been reduced from 3.596 down to 0.903 (achieved by increasing ϕ from 0.64 to 0.887). As in Figure 2.1 each line on the figure represents the spread of the various stages of infectivity among the addict population, and the total fraction of infected needles. It is clear from the figure that the disease dies out in all addicts and all needles and after about 110 years the Simple Model has almost

Figure 2.2: Simple Model when $R_0 > 1$

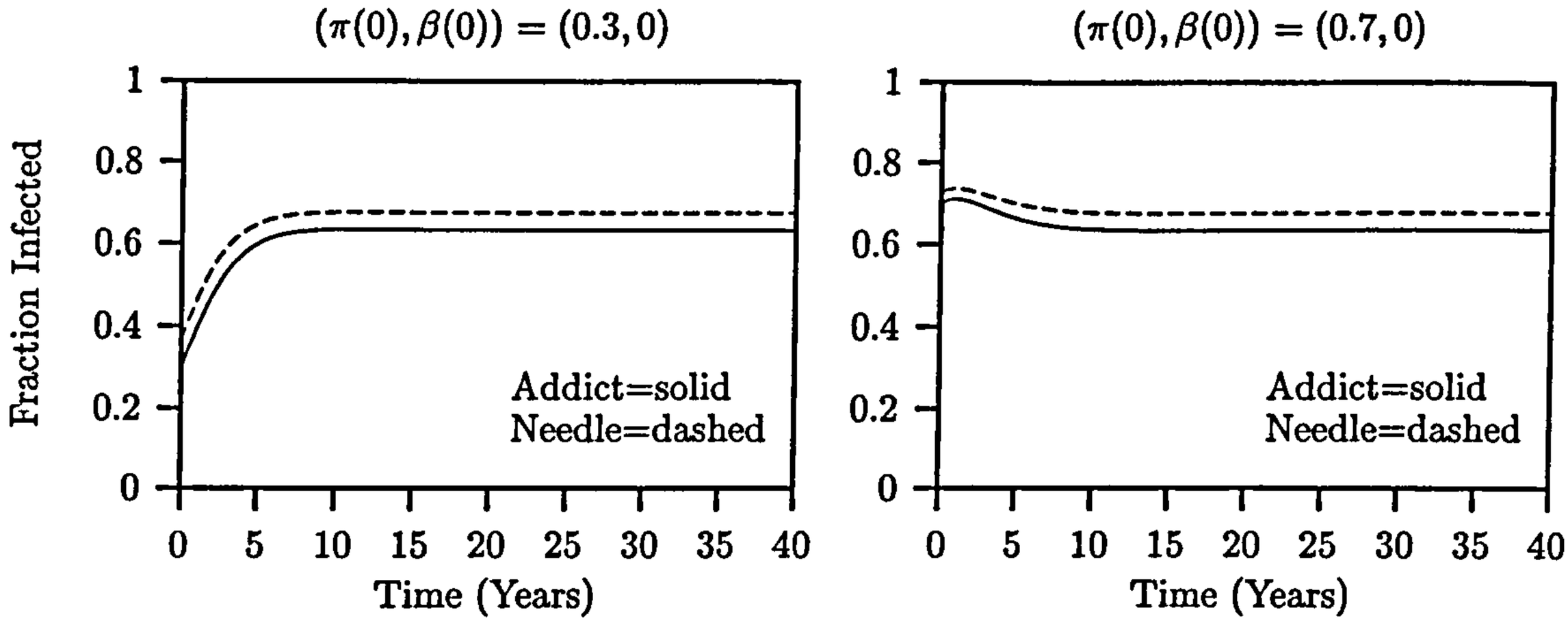


Figure 2.3: Simple Model when $R_0 > 1$

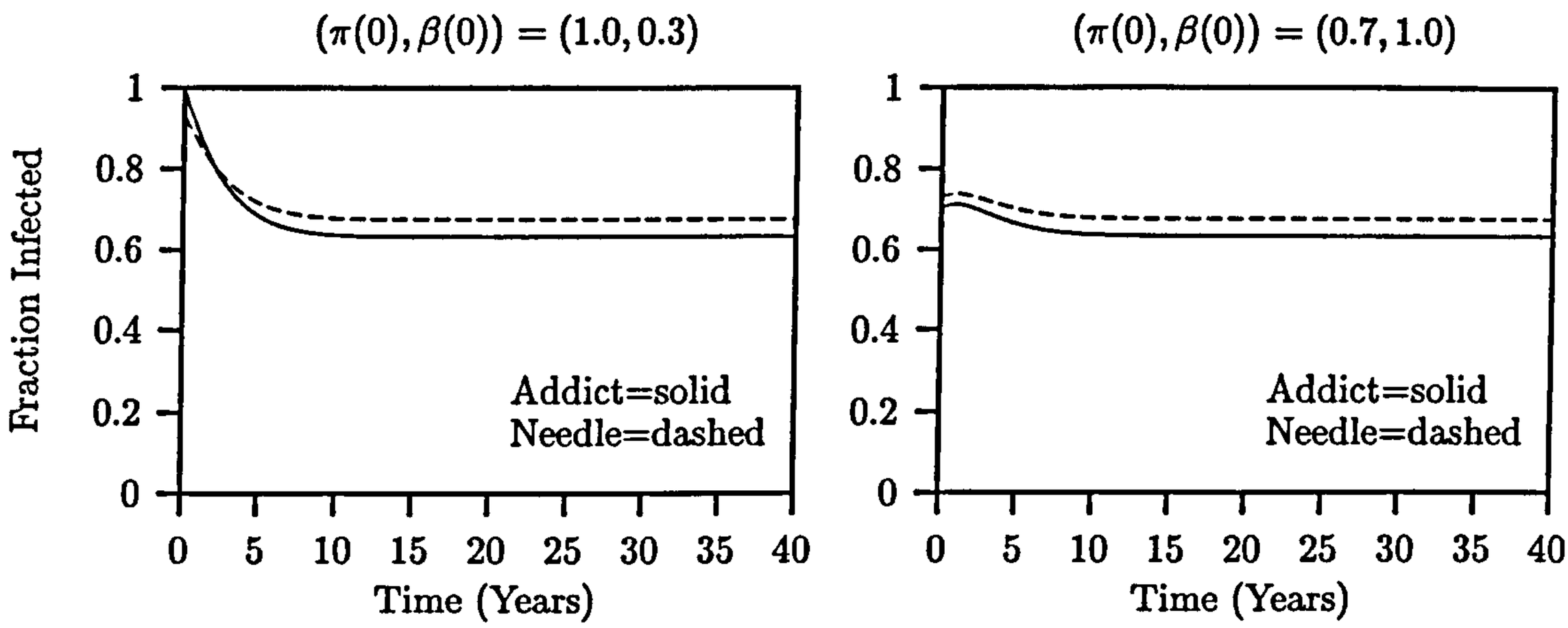


Figure 2.4: Simple Model when $R_0 > 1$

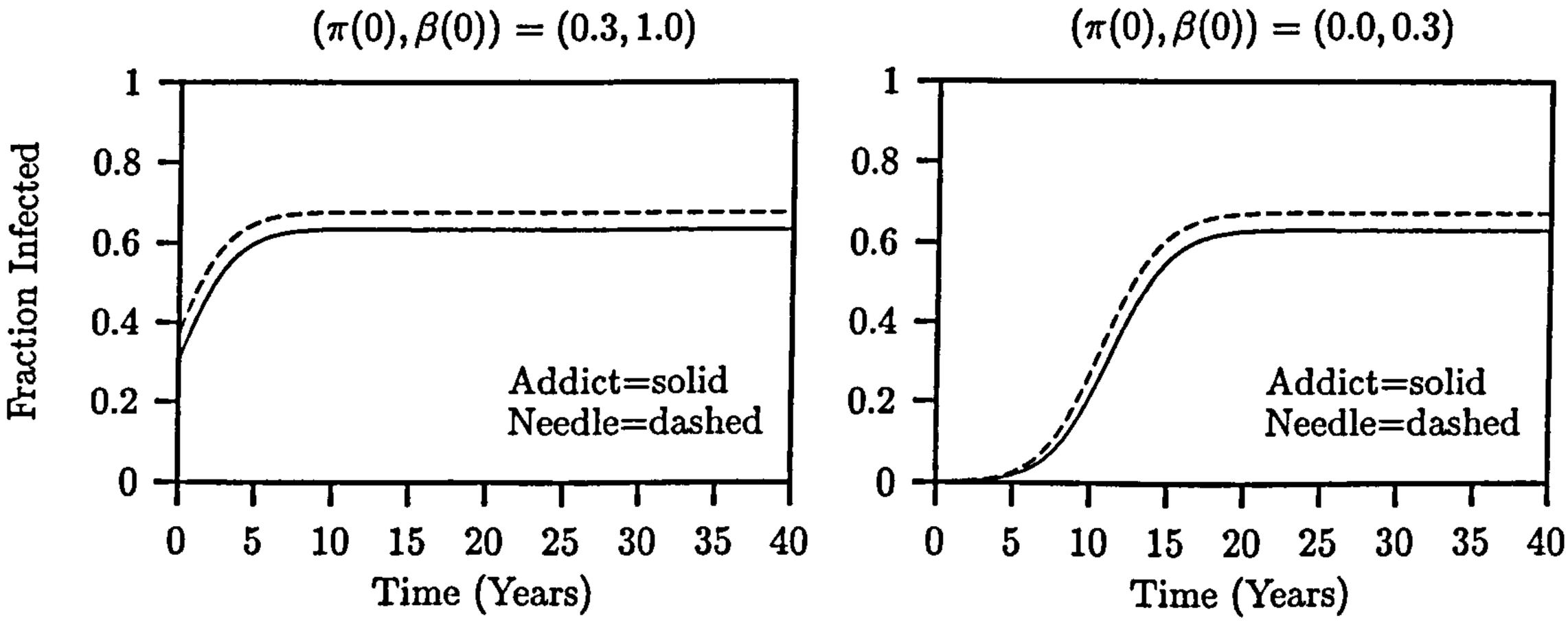
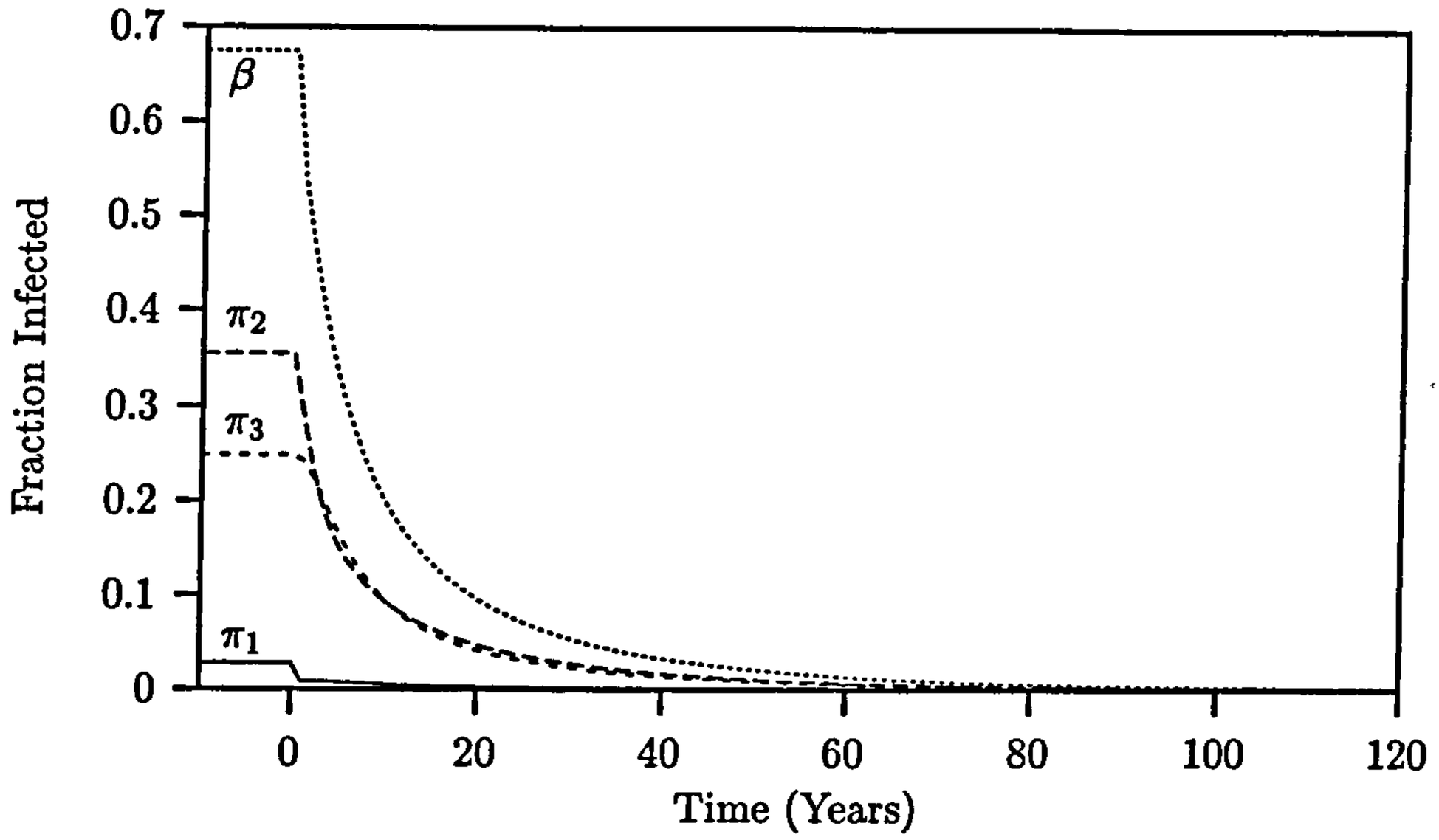


Figure 2.5: System Tends to Disease-Free Equilibrium when $R_0 < 1$



reached the disease-free equilibrium. Other simulations (not illustrated) show that for a variety of different parameter estimates and initial conditions the disease always dies out in both addicts and needles when $R_0 < 1$. This is consistent with Theorem 2.2.

2.7 Summary of Results for the Simple Model

We began this chapter by extending the modelling assumptions made by Kaplan and O'Keefe (1993) to incorporate three types of infectious addicts. We then derived a system of four differential equations which encapsulated the information in this extended set of assumptions. We next derived an expression for R_0 for this model by considering the impact that each type of infectious addict has on the infection process. Most of the remaining part of the chapter was devoted to examining the properties of the model equations, in particular the properties and behaviour of the equilibrium solutions. We showed analytically that the only condition necessary for the disease to die out in all addicts and all needles is $R_0 \leq 1$. Similarly we showed analytically that if initially present the only condition necessary for disease to persist among the population (for all time) is $R_0 > 1$. Moreover we demonstrated that the endemic equilibrium in our model is locally stable when $R_0 > 1$ and derived sufficient conditions for global stability of this equilibrium solution. We finally illustrated the dynamic and long term behaviour of our model graphically using a small number of numerical simulations.

The model discussed in this chapter allows addicts to progress through three different infectious stages. Addicts move from Acute Infectivity through to Asymptomatic and Pre-AIDS Infectivity according to a Poisson process. The model assumes that all addicts leave a needle with the same level of infectivity. This is a difficult assumption to check biologically as it is very difficult to measure the infectivity of an infectious needle. However a similar assumption was made by Peterson et al. (1990) and appears to have been made by Seitz and Müller (1994). Therefore whilst our three stage model allows addicts to progress through three different stages of infectivity the basic infection mechanism (through injecting with a contaminated needle) in this model is the same as in the Kaplan and O'Keefe Model.

It is not clear whether the assumption that all needles are equally infectious (with infectivity denoted by the single transmission parameter α) is biologically justified. It may be the case that infectivity is proportional to the level of virus in the blood. This would suggest that it may be more appropriate to divide the population of needles into three infectious classes corresponding to the infectious state of the last addict to use the needle. Thus we would allow addicts in stage one infectivity to leave a needle so that the probability of transmission of HIV from this needle is α_1 (say), similarly addicts in stage two or three infectivity would leave a needle so that the probability of transmission of HIV from this needle is α_2 and α_3 respectively, where α_1, α_2 and α_3 are proportional to the viral load of HIV in the blood of addicts in the respective infectious stages. We consider viral load as denoting the "amount of virus" present, for example we consider the viral load of a needle as representing the amount of virus resident in the needle and assume that the infectivity of a needle is proportional to this. Modelling this situation requires a more complex model with three classes of infectious needles and importantly we now require information relating to how addicts and needles of different classes interact with each other. For example we need to know which infectious class a needle will be left in after use by an addict who is currently acutely infectious. Obviously the outcome of this event will depend on the infectivity of the needle prior to use as well as that of the current user. These addict-needle interaction assumptions are very difficult to assess and there are very few or no empirical data to support any assumptions. In single stage infectivity the "flushing" parameter θ represents an assumption relating to how addicts and needles interact (specifically interactions of infectious needles with uninfectious addicts). Hence we need to generalise the "flushing" parameter in our addict-needle interaction assumptions. This is a difficult problem and we use three

separate models to investigate this. The following chapter deals with the first of these models.

Chapter 3

The Optimistic Model

3.1 Introduction

In this chapter we develop the first of three models which incorporate three stages of infectivity in both addicts *and* needles. It is natural to assume that allowing addicts to progress through three stages of infectivity should also carry implications for the infectivity of needles circulating among the addict population. We first define three types of infectious needles, we then develop a model which incorporates these three types of needles. This model is constructed using a particular set of assumptions concerning how addicts and needles of different levels of infectivity interact with each other. It is very difficult to realistically assess addict-needle interactions, in response to this problem our model deliberately uses assumptions which may be more optimistic than might reasonably be expected. Once we have derived the model we then compute an expression for the basic reproductive number and conduct an equilibrium and stability analysis in a similar manner to the Simple Model in the previous chapter. The chapter concludes with a brief summary of the main findings.

3.2 Infectious Needle Definitions

We wish to construct a model which allows addicts to progress through three distinct stages of infectivity prior to the onset of full blown AIDS. By allowing addicts to exist in three stages of infectivity it is obviously more realistic to also allow needles to possess different levels of infectivity (since it is the blood in addicts which determines the infectiousness of needles). A natural way of splitting up the class of infectious needles is to use three classes, where each class corresponds to the infectivity of a type

of infectious addict. We now discuss how we shall define a state one, state two and state three infectious needle.

Intuitively, if a previously unused[1] needle is used by an addict in stage one infectivity then the needle will be left contaminated with the HIV virus, the concentration of which is dependent on the viral load in the blood of the addict. Similarly if a stage two or stage three infectious addict injects with a previously unused needle then each addict will leave a different amount of HIV virus in the needle. Let us define a state one infectious needle as a previously unused needle which has been used once by an addict in stage one infectivity. Similarly we define a state two and state three infectious needle as a previously unused needle which has been used once by an addict in stage two or stage three infectivity respectively. These definitions are sensible, the only caveat being that we have assumed that all addicts inject in an identical manner, this is important since the method of injecting could affect how much of the addicts' blood will remain in the needle. To the best of our knowledge this assumption has been made in all previous models involving needle sharing intravenous drug users.

We have defined what we shall refer to as a state one, state two and state three infectious needle. However whilst our definition is sensible it is not of practical use in a model of needle sharing since a needle will most probably be used more than once before being removed from the population (therefore it is unlikely that needles will be devoid of fluid prior to use as in our definition). We need to modify this definition to provide something more realistic for use in our models. As a first step towards a more practical definition we extend the definition to take into account needles which have previously been used but are still uncontaminated. Hence we define a state one, state two and state three infectious needle as *any* uncontaminated needle which has been used by an addict in stage one, stage two or stage three infectivity respectively.

Intuitively, a previously unused needle used for the first time by a susceptible addict will leave an amount of uncontaminated blood in the needle after use. Due to this extra uncontaminated blood it is possible that the HIV viral load left in the needle by the next infectious user would differ from the case where the infectious addict injected with a previously unused needle. However it is also possible that the infectious addict flushes the needle, this would presumably leave the HIV viral load in the needle similar to the case where the needle was new and unused. It is not clear which of these situations is more plausible, and we are not aware of any data available to assist with this. We

[1] we assume that an unused needle is a needle devoid of any fluid.

assume that the difference between the HIV viral load left by an infectious addict in an unused needle and in a used but uncontaminated needle is sufficiently small to be ignored, and in both cases the needle is left with the same viral load as in the blood of the infectious addict.

3.3 Addict-Needle Interaction Assumptions

We have defined the three types of infectious needles that we wish to incorporate into our model. Now we must decide on how the four types of addicts and four types of needles in our model interact (we have three types of infectious addicts and needles and one type of uninfectious addict and needle). More exactly we must specify for $i, j, k = 0, 1, 2, 3$ what fraction p_{ijk} of needles initially in infectious state i are left in infectious state k after use by an addict in infectious stage j. This gives us sixty-four potential needle-addict interactions to specify. However for sixteen of these cases the answer is obvious. If $i = j$ then the initial infectious state of the needle is equal to the infectious stage of the addict, hence the final infectious state of the needle must obviously be $k = i = j$. This leaves us requiring to specify forty-eight needle-addict interactions. To give a concrete example suppose that a state one needle is used by a stage three addict, we wish to determine which infectious class the needle enters after the addict has injected with it. Note that the outcome of this is not necessarily the same as the outcome of a state three needle and a stage one addict interaction. The reason for this is that the volume of addict's blood which is drawn into a needle is not necessarily the same as the volume of residual blood already left in the needle from the previous user, hence the HIV viral load left in the needle may differ in each case.

It is very difficult to determine the outcome of many of these forty-eight addict-needle interactions. It is clear that important factors in the outcome of each interaction are differences in HIV viral load between the different infectious stages, the volume of addict's blood which is drawn into a needle and the volume of blood already in the needle from the previous user. Unfortunately there are no empirical data to aid with the problem of estimating these p_{ijk} probabilities. Research has been carried out to ascertain the relative HIV viral load in human blood during each stage of infectivity (Ward et al., 1987), however to the best of our knowledge this is the extent of the data. Peterson et al. (1990) assume that viral loads in stages one, two and three are in the ratio 5:1:3. While this data is useful it does not assist directly in determining the

outcome of any of the addict needle interactions as we can only guess at the difference between the volume of blood drawn into a needle and the volume of residual blood left behind in the needle after an addict has used it. We could make an educated guess as to the most sensible outcome of each addict-needle interaction, however it is probably inevitable that whatever assumptions we make these will be inaccurate, although just how much so is difficult to say.

Rather than try to guess the most realistic addict-needle interaction assumptions based on little hard evidence we can instead choose a set of addict-needle interaction assumptions for the remaining unspecified p_{ijk} probabilities which should be more optimistic than would reasonably be expected. By optimistic we mean that the incidence rate and prevalence of the disease should be less. In this way we can establish a lower bound for the prevalence of HIV among intravenous drug users under the assumption of a three stage infectious period.

There is a more general way of dealing with the problem of deciding needle-addict interactions. Instead of adopting a particular set of assumptions we can assign an unspecified probability p_{ijk} to the outcome of each addict-needle interaction. This approach leads us to a single, much more complex, model which is discussed in Chapter 5.

3.4 Optimistic Addict-Needle Interaction Assumptions

We expect that when a needle is used by an addict then the needle is left at a level of infectiousness somewhere between its initial level and the level of the addict who last used it. For example if a state one infectious needle is used by a stage two infectious addict then it is reasonable to expect that the needle is left with a viral load somewhere between that corresponding to state one and state two, so it must be left in infectious state one, two or three, it cannot be left uninfected.

We now examine a set of addict-needle interaction assumptions which give rise to a model which is deliberately the most optimistic possible. This is useful because we can hopefully use it to obtain bounds for the fraction of needles and addicts infected. We assume that all needles take on the relative characteristics of the last user, for example a state one infectious needle used by a stage two infectious addict will become a state two infectious needle, a state three needle used by a susceptible addict will become uncontaminated and so on. This will be appropriate if almost all of the blood initially

contained in the syringe is injected into the user and replaced by the infectious blood of the injecting addict.

The mechanism by which drug users inject heroin is as follows: the drug is first dissolved in water and drawn into a syringe, this solution is then injected into the addict. There will be some residual blood left in the syringe after the initial injection and further blood from the addict may be drawn back into the syringe and re-injected in order to get the full benefit of the drug (Blower et al., 1991). This could leave contaminated blood in the syringe (Samuels et al., 1992). As it is not unreasonable to suppose that the volume of blood transferred from the addict into the syringe during injection is much greater than the amount of residual infected blood from the last addict it is possible that our optimistic assumptions may be reasonably close to reality.

In the Simple Model it was assumed that needles could be flushed with probability θ. From the expression for the endemic equilibrium solution for this model it is obvious that if all else is held constant then increasing θ from zero to one lowers the endemic level of HIV in addicts. In other words flushing is a beneficial action in the model with a single type of infectious needle. Our optimistic assumptions generalise the case where $\theta = 1$ as a basis for our lower bound three stage model, and are a similar idea to assuming that all addicts flush all needles with probability one.

3.5 Model Derivation

We wish to expand the model defined by equations (2.1)-(2.4) in Chapter 2 to incorporate three types of infectious needle. We do not alter the behaviour of the addicts and as such the addict equations are the same except that now

$$\frac{d\pi_1}{dt} = \Big(1 - \sum_{i=1}^{3} \pi_i\Big)\lambda(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1-\phi) - (\mu + \delta_1)\pi_1.$$

We have replaced $\beta\alpha$ with $\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$ to incorporate the fact that addicts can now be infected by three different types of infectious needle, each with its own HIV transmission probability. The probability of becoming infected with HIV in a single injection from a state one needle is α_1, and similarly the probability of becoming infected with HIV in a single injection from a state two or state three needle is α_2 or α_3 respectively. We now move on to the three equations which describe the behaviour of the needles.

The number of infected state one needles at time $t + \Delta t$

$$
\begin{aligned}
= \quad & \{\text{number of state one infectious needles at time } t\} \\
& + \{(\text{number of non state one needles at time } t) \\
& \times (\text{fraction of needles used by stage one addicts in } [t, t+\Delta t))\} \\
& - \{(\text{number of state one infected needles at time } t) \\
& \times (\text{fraction of needles used by non stage one addicts in } [t, t+\Delta t))\} \\
& - \{\text{number of state one infectious needles exchanged in } [t, t+\Delta t)\}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
m\beta_1(t+\Delta t) &= m\beta_1(t) + m\lambda\Delta t\gamma\pi_1(t)(\beta_2(t) + \beta_3(t) + (1 - \beta_1(t) - \beta_2(t) - \beta_3(t))) \\
&\quad - m\lambda\Delta t\gamma\beta_1(\pi_2(t) + \pi_3(t) + (1 - \pi_1(t) - \pi_2(t) - \pi_3(t))) - m\beta_1(t)\tau\Delta t \\
&\quad + o(\Delta t), \\
&= m\beta_1(t) + m\lambda\Delta t\gamma(\pi_1 - \beta_1) - m\beta_1\tau\Delta t + o(\Delta t).
\end{aligned}
$$

Subtracting $m\beta_1(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\beta_1}{dt} = \lambda\gamma(\pi_1 - \beta_1) - \beta_1\tau.$$

The number of infected state two needles at time $t + \Delta t$

$$
\begin{aligned}
= \quad & \{\text{number of state two infectious needles at time } t\} \\
& + \{(\text{number of non state two needles at time } t) \\
& \times (\text{fraction of needles used by stage two addicts in } [t, t+\Delta t))\} \\
& - \{(\text{number of state two infected needles at time } t) \\
& \times (\text{fraction of needles used by non stage two addicts in } [t, t+\Delta t))\} \\
& - \{\text{number of state two infectious needles exchanged in } [t, t+\Delta t)\}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
m\beta_2(t+\Delta t) &= m\beta_2(t) + m\lambda\Delta t\gamma\pi_2(t)(\beta_1(t) + \beta_3(t) + (1 - \beta_1(t) - \beta_2(t) - \beta_3(t))) \\
&\quad - m\lambda\Delta t\gamma\beta_2(\pi_1(t) + \pi_3(t) + (1 - \pi_1(t) - \pi_2(t) - \pi_3(t))) - m\beta_2(t)\tau\Delta t \\
&\quad + o(\Delta t), \\
&= m\beta_2(t) + m\lambda\Delta t\gamma(\pi_2 - \beta_2) - m\beta_2\tau\Delta t + o(\Delta t).
\end{aligned}
$$

Subtracting $m\beta_2(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\beta_2}{dt} = \lambda\gamma(\pi_2 - \beta_2) - \beta_2\tau.$$

The number of infected state three needles at time $t + \Delta t$

$$\begin{aligned} &= \{\text{number of state three infectious needles at time } t\} \\ &+\{(\text{number of non state three needles at time } t) \\ &\times(\text{fraction of needles used by stage three addicts in } [t, t+\Delta t))\} \\ &-\{(\text{number of state three infected needles at time } t) \\ &\times(\text{fraction of needles used by non stage three addicts in } [t, t+\Delta t))\} \\ &-\{\text{number of state three infectious needles exchanged in } [t, t+\Delta t)\}. \end{aligned}$$

Thus

$$\begin{aligned} m\beta_3(t+\Delta t) &= m\beta_3(t) + m\lambda\Delta t\gamma\pi_3(t)(\beta_1(t) + \beta_2(t) + (1 - \beta_1(t) - \beta_2(t) - \beta_3(t))) \\ &\quad -m\lambda\Delta t\gamma\beta_3(\pi_1(t) + \pi_2(t) + (1 - \pi_1(t) - \pi_2(t) - \pi_3(t))) - m\beta_3(t)\tau\Delta t \\ &\quad +o(\Delta t), \\ &= m\beta_3(t) + m\lambda\Delta t\gamma(\pi_3 - \beta_3) - m\beta_3\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m\beta_3(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\beta_3}{dt} = \lambda\gamma(\pi_3 - \beta_3) - \beta_3\tau.$$

Hence the system of differential equations which describes the spread of the disease is:

$$\frac{d\pi_1}{dt} = \Big(1 - \sum_{i=1}^{3}\pi_i\Big)\lambda(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1-\phi) - (\mu + \delta_1)\pi_1, \tag{3.1}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu + \delta_2)\pi_2, \tag{3.2}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu + \delta_3)\pi_3, \tag{3.3}$$

$$\frac{d\beta_1}{dt} = \lambda\gamma(\pi_1 - \beta_1) - \beta_1\tau, \tag{3.4}$$

$$\frac{d\beta_2}{dt} = \lambda\gamma(\pi_2 - \beta_2) - \beta_2\tau, \tag{3.5}$$

and

$$\frac{d\beta_3}{dt} = \lambda\gamma(\pi_3 - \beta_3) - \beta_3\tau, \tag{3.6}$$

with suitable initial conditions: $\pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$, $\beta_1(0)$, $\beta_2(0)$ and $\beta_3(0)$ where $\pi_i(0) \geq 0$ and $\beta_j(0) \geq 0$, for $i, j = 1, 2, 3$, $\pi_1(0) + \pi_2(0) + \pi_3(0) \leq 1$ and $\beta_1(0) + \beta_2(0) + \beta_3(0) \leq 1$. The needle equations do not feature the parameter ϕ, the probability that an addict successfully cleans a needle prior to use. The reason for this is that under the assumption of full flushing we have that all the contents of a needle prior to use are removed by injection (into the addict) during the injection process. Hence the state of a needle after use is determined purely by the HIV viral load of the residual blood left by the current user. Thus cleaning a needle prior to use has no effect on the state of the needle after use.

3.6 The Basic Reproductive Number

As in Section 2.4 we are interested in deriving an explicit expression for the basic reproductive number. This time we are interested in the value of R_0 based on the model defined by eqns (3.1)-(3.6).

As usual we consider a single newly infectious addict entering a totally susceptible population of addicts and needles. The infectious addict is initially in stage one infectivity. Following a similar method as in Section 2.4 we first determine the expected number of needles an addict will infect during each infectious stage of his or her infectious lifetime. This time we have three types of infectious needle, the assumptions in our model imply that when an addict is in stage one he or she leaves needles in stage one infectivity, when in stage two, needles are left in stage two infectivity and when in stage three, needles are left in stage three infectivity. We know that on average an addict infects

$$\frac{\lambda}{\mu + \delta_1}$$

needles during his or her entire stage one lifetime, and

$$\frac{\lambda\delta_1}{(\mu + \delta_1)(\mu + \delta_2)}$$

needles during his or her entire stage two lifetime, and

$$\frac{\lambda\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}$$

needles during his or her entire stage three lifetime. Hence we now know how many of each type of needle an addict infects during his or her entire infectious lifetime. We now wish to determine how many infections will be caused by each type of needle until it is rendered virus free. By treating each type of needle separately we can proceed in the same manner as for the Simple Model. Firstly we consider a single stage one infectious needle, using the same notation as in Section 2.4 and where $E_1 =$ E(addicts infected by a single state one needle until it is flushed), we have that

$$\begin{aligned} E_1 &= \frac{(1-\phi)\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected and needle flushed)} \\ &\qquad +\text{P(sus. addict infected and needle not flushed)}(E_1+1) \\ &\qquad +\text{P(sus. addict not infected and needle not flushed)}E_1\Big], \\ &= \frac{(1-\phi)\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected) + P(needle not flushed)}E_1\Big]. \end{aligned}$$

However for the Optimistic Model we have assumed that needles are *always* flushed, so solving for E_1 simply gives

$$E_1 = \frac{(1-\phi)\alpha_1}{\hat{\tau}+1},$$

where $\hat{\tau} = \tau/\lambda\gamma$. Following an identical argument for state two and state three infectious needles we find that

$$E_2 = \frac{(1-\phi)\alpha_2}{\hat{\tau}+1},$$

$$\text{and} \quad E_3 = \frac{(1-\phi)\alpha_3}{\hat{\tau}+1}.$$

We now know the expected number of each type of infected needle an addict creates during his or her entire infectious lifetime, and we know the expected number of addicts each of these types of infectious needle infects. Multiplying these two values gives us,

$$R_0 = \frac{\lambda(1-\phi)}{(\mu+\delta_1)(\hat{\tau}+1)}\left[\alpha_1+\frac{\alpha_2\delta_1}{\mu+\delta_2}+\frac{\alpha_3\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right]. \tag{3.7}$$

We now move on to exploring the properties of the Optimistic Model in a similar manner to the Simple Model in the previous chapter. As before we are interested in the importance of the value of R_0 in determining the long term behaviour of our model.

3.7 Analytical Results

We now investigate analytically the properties of the Optimistic Model. The mathematical results in this section share a number of things in common with the results for the Simple Model. In particular the Optimistic and Simple Models share the same addict equations and this allows us to use certain results from Section 2.5 directly. As in the Simple Model we assume that all model parameters except ϕ are strictly positive and that ϕ is strictly less than unity. In addition we assume that $\alpha_1 \geq \alpha_3 \geq \alpha_2$, this implies that needles in state one infectivity are more infectious than those in state three infectivity which are more infectious than those in state two infectivity. We are primarily interested in the behaviour of the equilibrium solutions and in particular whether $R_0 \leq 1$ is still a necessary and sufficient condition for the disease to die out in all addicts and needles, and similarly whether $R_0 > 1$ is still a necessary and sufficient condition for the disease to persist among the population if it is present initially.

Theorem 3.1 *If $R_0 \leq 1$ the system of equations (3.1)-(3.6) has a unique equilibrium solution where the disease has died out in both addicts and needles. If $R_0 > 1$ then there is still the equilibrium where the disease has died out, however there is also a unique endemic equilibrium.*

Proof.

Let $\pi_1^\star$, $\pi_2^\star$, $\pi_3^\star$, $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ denote the respective equilibrium proportions of infected addicts and infected needles in infectious stages one, two and three. Let $\pi = \pi_1 + \pi_2 + \pi_3$, the total proportion of infected addicts, $\beta = \beta_1 + \beta_2 + \beta_3$, the total proportion of infected needles, and $L = 1 + \delta_1/(\mu+\delta_2) + (\delta_1\delta_2)/((\mu+\delta_2)(\mu+\delta_3))$. Using eqns (3.1)-(3.3) we can write $\pi_1^\star$, $\pi_2^\star$ and $\pi_3^\star$ in terms of $\pi^\star$ as follows,

$$\pi_1^\star = \frac{\pi^\star}{L},$$

$$\pi_2^\star = \frac{\delta_1}{(\mu+\delta_2)}\frac{\pi^\star}{L},$$

and

$$\pi_3^\star = \frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\frac{\pi^\star}{L}.$$

From eqns (3.4)-(3.6) we find that

$$\beta_1^\star = \frac{\pi^\star}{L(1+\hat{\tau})}, \tag{3.8}$$

$$\beta_2^\star = \frac{\pi^\star\delta_1}{L(1+\hat{\tau})(\mu+\delta_2)}, \tag{3.9}$$

and $$\beta_3^\star = \frac{\pi^\star \delta_1 \delta_2}{L(1+\hat{\tau})(\mu+\delta_2)(\mu+\delta_3)}. \tag{3.10}$$

Substituting the expressions for $\beta_1^\star, \beta_2^\star$ and $\beta_3^\star$ into eqn (3.1) and substituting $\pi^\star/L$ for $\pi_1^\star$ gives

$$(1-\pi^\star)\lambda(1-\phi)\frac{E}{L} = \frac{(\mu+\delta_1)}{L},$$

where

$$E = \frac{\alpha_1}{1+\hat{\tau}} + \frac{\alpha_2\delta_1}{(1+\hat{\tau})(\mu+\delta_2)} + \frac{\alpha_3\delta_1\delta_2}{(1+\hat{\tau})(\mu+\delta_2)(\mu+\delta_3)}.$$

Hence

$$\pi^\star = 1 - \frac{\mu+\delta_1}{\lambda(1-\phi)E},$$

and $$\beta^\star = \frac{\pi^\star}{1+\hat{\tau}}, \qquad \text{using eqns (3.4)-(3.6).}$$

Using the expression of R_0 from eqn (3.7) we find that

$$(\pi^\star, \beta^\star) = \left(\frac{R_0-1}{R_0}, \frac{R_0-1}{R_0(1+\hat{\tau})}\right). \tag{3.11}$$

As in the proof of Theorem 2.1 there can only be two equilibrium solutions, the disease-free solution and a strictly positive solution. From eqn (3.11) it is obvious that if $R_0 \leq 1$ then the only solution is the disease-free solution, and if $R_0 > 1$ then we have a unique endemic solution. This completes the proof of Theorem 3.1. •

Theorem 3.2 *If $R_0 \leq 1$ then whatever the initial state the disease will die out in both addicts and needles.*

Proof.

This proof is very similar to that of Theorem 2.2. Eqns (3.2)-(3.3) are the same as eqns (2.2)-(2.3) thus we can use Lemma 2.1 and Corollary 2.1 directly. The form of eqns (3.4)-(3.6) is very similar to eqns (2.2)-(2.3), hence we have directly that if $\beta_i^\infty = \limsup_{t\to\infty} \beta_i(t)$ for $i = 1,2,3$ then

$$\beta_i^\infty \leq \frac{\pi_i^\infty}{1+\hat{\tau}},$$

for $i = 1,2,3$. Using Lemma 2.1 and Corollary 2.1 we find that

$$\beta_2^\infty \leq \frac{\delta_1\pi_1^\infty}{(\mu+\delta_2)(1+\hat{\tau})},$$

and $$\beta_3^\infty \leq \frac{\delta_1\delta_2\pi_1^\infty}{(\mu+\delta_2)(\mu+\delta_3)(1+\hat{\tau})}.$$

Now suppose that $\pi_1^\infty > 0$. Given $\epsilon > 0$,

$$\begin{aligned}
\frac{d\pi_1}{dt} &\leq (1-\pi_1)\lambda(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)(1-\phi)-(\mu+\delta_1)\pi_1, \\
&\leq (1-\pi_1)\lambda(1-\phi)(\beta_1^\infty\alpha_1+\beta_2^\infty\alpha_2+\beta_3^\infty\alpha_3+\epsilon)-(\mu+\delta_1)\pi_1, \qquad \forall t \geq t_1(\epsilon) \\
&\leq (1-\pi_1)\frac{\lambda(1-\phi)}{1+\hat{\tau}}\left[\alpha_1+\frac{\delta_1\alpha_2}{\mu+\delta_2}+\frac{\delta_1\delta_2\alpha_3}{(\mu+\delta_2)(\mu+\delta_3)}+\epsilon_1\right]\pi_1^\infty-(\mu+\delta_1)\pi_1, \\
&\qquad\qquad \text{where } \epsilon_1 = \frac{\epsilon(1+\hat{\tau})}{\pi_1^\infty}, \\
&\leq (1-\pi_1)(\mu+\delta_1)(R_0+\epsilon_2)\pi_1^\infty-(\mu+\delta_1)\pi_1, \qquad \text{where } \epsilon_2 = \frac{\lambda(1-\phi)\epsilon_1}{(\mu+\delta_1)(1+\hat{\tau})}, \\
&\leq (\mu+\delta_1)[(R_0+\epsilon_2)\pi_1^\infty-(1+R_0\pi_1^\infty)\pi_1].
\end{aligned}$$

The result now follows directly using the latter part of the proof of Theorem 2.2. Once we have established that $\pi_1^\infty = 0$ it follows immediately that π_2^∞, π_3^∞, β_1^∞, β_2^∞ and β_3^∞ are all zero and then that $\pi_1(t)$, $\pi_2(t)$, $\pi_3(t)$, $\beta_1(t)$, $\beta_2(t)$ and $\beta_3(t)$ all tend to zero as $t \to \infty$.•

Theorem 3.3 *If $R_0 > 1$ then there is still the equilibrium where the disease has died out and this equilibrium is unstable.*

Proof.

Consider the linearised system of eqns (3.1)-(3.6), evaluated at the disease-free equilibrium. This system can be represented in matrix form as

$$\frac{d\mathbf{x}}{dt} = \mathbf{J}\mathbf{x},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$ and

$$\mathbf{J} = \begin{bmatrix}
-(\mu+\delta_1) & 0 & 0 & \lambda\alpha_1(1-\phi) & \lambda\alpha_2(1-\phi) & \lambda\alpha_3(1-\phi) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\
\lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 & 0 \\
0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 \\
0 & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau)
\end{bmatrix}.$$

We wish to show that at least one eigenvalue of J has a strictly positive real part. Using the Routh-Hurwitz conditions it is sufficient to show that the constant term, a_6, in the characteristic equation of J,

$$\omega^6 + a_1\omega^5 + a_2\omega^4 + a_3\omega^3 + a_4\omega^2 + a_5\omega + a_6 = 0$$

is strictly negative. It is straightforward to show that

$$\begin{aligned} a_6 &= (\lambda\gamma+\tau)(\lambda\gamma+\tau)(\lambda\gamma+\tau)(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3) \\ &\quad -\lambda\alpha_3(1-\phi)\lambda\gamma(\lambda\gamma+\tau)(\lambda\gamma+\tau)\delta_1\delta_2 \\ &\quad -\lambda\alpha_2(1-\phi)(\lambda\gamma+\tau)(\lambda\gamma+\tau)\lambda\gamma\delta_1(\mu+\delta_3) \\ &\quad -\lambda\gamma(\lambda\gamma+\tau)(\lambda\gamma+\tau)(\mu+\delta_2)(\mu+\delta_3)\lambda\alpha_1(1-\phi), \\ &= (\lambda\gamma+\tau)^3(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3) \\ &\quad \times\left[1-\frac{\lambda^2\gamma\alpha_1(1-\phi)(\mu+\delta_2)(\mu+\delta_3)+\lambda^2\gamma\alpha_2(1-\phi)\delta_1(\mu+\delta_3)}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma+\tau)}\right. \\ &\quad \left.-\frac{\lambda^2\gamma\alpha_3(1-\phi)\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma+\tau)}\right]. \end{aligned}$$

Substituting in the expression for R_0 in eqn (3.7) we find that

$$a_6 = (\lambda\gamma+\tau)^3(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(1-R_0),$$

hence if $R_0 > 1$ then $a_6 < 0$ and the result follows.•

Theorem 3.4 *If $R_0 > 1$ and either $\pi(0) > 0$ or $\beta(0) > 0$ then there exists a fixed $\epsilon > 0$ depending only on the model parameters and not the initial conditions such that for some $\eta > 0$*

$$\pi_1 \geq \epsilon\pi_1^\star,\ \pi_2 \geq \epsilon\pi_2^\star,\ \pi_3 \geq \epsilon\pi_3^\star,\ \beta_1 \geq \epsilon\beta_1^\star,\ \beta_2 \geq \epsilon\beta_2^\star,\ \beta_3 \geq \epsilon\beta_3^\star, \quad \forall t \geq \eta. \tag{3.12}$$

As in the case of the Simple Model in the previous chapter the proof of this theorem requires a number of supporting arguments. The method of the proof is similar to that for the Simple Model, the main difference is the extra work involved in dealing with six rather than just four equations which describe the disease dynamics. This result follows the same intuitive argument as for the Simple Model, the two main parts of which are that π_1 is the dominant component of the system and that the disease-free equilibrium is unstable for this model when $R_0 > 1$. The Optimistic Model has the same equations for $d\pi_2/dt$ and $d\pi_3/dt$ as the Simple Model, hence we can use Lemma 2.2 and Corollary 2.3 directly, thus we have that

$$\pi_{2,\infty} \geq \frac{\delta_1\pi_{1,\infty}}{\mu+\delta_2}, \tag{3.13}$$

and

$$\pi_{3,\infty} \geq \frac{\delta_1\delta_2\pi_{1,\infty}}{(\mu+\delta_2)(\mu+\delta_3)}. \tag{3.14}$$

From eqns (3.4)-(3.6) we find that

$$\frac{d}{dt}\Big(\beta_1 \exp[(\lambda\gamma+\tau)t]\Big) = \pi_1\lambda\gamma\exp[(\lambda\gamma+\tau)t], \tag{3.15}$$

$$\frac{d}{dt}\Big(\beta_2 \exp[(\lambda\gamma+\tau)t]\Big) = \pi_2\lambda\gamma\exp[(\lambda\gamma+\tau)t], \tag{3.16}$$

and
$$\frac{d}{dt}\Big(\beta_3 \exp[(\lambda\gamma+\tau)t]\Big) = \pi_3\lambda\gamma\exp[(\lambda\gamma+\tau)t]. \tag{3.17}$$

It follows directly that

$$\beta_{1,\infty} \geq \frac{\lambda\gamma\pi_{1,\infty}}{\lambda\gamma+\tau}, \tag{3.18}$$

$$\beta_{2,\infty} \geq \frac{\lambda\gamma\pi_{2,\infty}}{\lambda\gamma+\tau} \geq \frac{\lambda\gamma\delta_1\pi_{1,\infty}}{(\lambda\gamma+\tau)(\mu+\delta_2)}, \tag{3.19}$$

and
$$\beta_{3,\infty} \geq \frac{\lambda\gamma\pi_{3,\infty}}{\lambda\gamma+\tau} \geq \frac{\lambda\gamma\delta_1\delta_2\pi_{1,\infty}}{(\lambda\gamma+\tau)(\mu+\delta_2)(\mu+\delta_3)}. \tag{3.20}$$

From the previous results it is sufficient to show that $\pi_{1,\infty} > 0$ since then given $\nu > 0$ there exists $\zeta(\nu)$ such that $\pi_1 \geq \pi_{1,\infty} - \nu$ for $t \geq \zeta(\nu)$. Hence to bound the trajectories away from zero it is sufficient to show that $\pi_{1\,\infty} > 0$. Choosing $\nu = \frac{1}{2}\pi_{1,\infty}$ gives $\pi_1 \geq \frac{1}{2}\pi_{1,\infty}$ for $t \geq \zeta(\nu)$. Inequalities (3.13)-(3.14) and (3.18)-(3.20) imply that $\pi_{2,\infty}$, $\pi_{3,\infty}$, $\beta_{1,\infty}$, $\beta_{2,\infty}$ and $\beta_{3,\infty}$ are all strictly positive when $\pi_{1,\infty} > 0$ and hence (3.12) holds.

Lemma 3.1 *Provided that at least one of* $\pi_1(t)$, $\pi_2(t)$, $\pi_3(t)$, $\beta_1(t)$, $\beta_2(t)$ *and* $\beta_3(t)$ *is strictly positive at* $t = 0$ *then* $\pi_1(\Delta t) > 0$, $\pi_2(\Delta t) > 0$, $\pi_3(\Delta t) > 0$, $\beta_1(\Delta t) > 0$, $\beta_2(\Delta t) > 0$ *and* $\beta_3(\Delta t) > 0$ *for* Δt *small and positive.*

Proof.

We need to consider four separate initial conditions:

1. Suppose that $\beta(0) = 0$. Hence $\pi(0) > 0$. Using a Taylor expansion about $t = 0$ and eqns (3.1)-(3.6) we find that

$$\beta(\Delta t) = \pi(0)\lambda\gamma\Delta t + o(\Delta t) > 0,$$

and
$$\pi(\Delta t) = \pi(0) - (\mu\pi(0) + \delta_3\pi_3(0))\Delta t + o(\Delta t) > 0$$

(for Δt small and positive).

Let $\psi = 1 - \pi$, hence

$$\frac{d\psi}{dt} = -\psi\lambda\hat{\beta}(1-\phi) + \mu(1-\psi) + \pi_3\delta_3,$$

where $\hat{\beta} = \beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$. If $\pi(0) < 1$ then we have that $\psi(0) > 0$, hence $\psi(\Delta t) > 0$ for small enough Δt. If $\pi(0) = 1$ then $\psi(0) = 0$ and

$$\psi(\Delta t) \geq \mu\Delta t + o(\Delta t) > 0 \qquad \text{(for } \Delta t \text{ small and positive).}$$

Hence by choosing Δt small enough and starting at $t = \Delta t$ we can assume that (if necessary) $\pi(0) > 0$, $\psi(0) > 0$ and $\beta(0) > 0$. If $\pi_1(0) = 0$ then $\pi_1(\Delta t) = \psi(0)\lambda\hat{\beta}(0)(1-\phi)\Delta t + o(\Delta t) > 0$, hence we can also assume that $\pi_1(0) > 0$. If $\beta_1(0) = 0$ then $\beta_1(\Delta t) = \lambda\gamma\pi_1(0)\Delta t + o(\Delta t) > 0$ hence we can assume that $\beta_1(0) > 0$. If $\pi_2(0) = 0$ then $\pi_2(\Delta t) = \delta_1\pi_1(0)\Delta t + o(\Delta t) > 0$, hence we can also assume that $\pi_2(0) > 0$. If $\beta_2(0) = 0$ then $\beta_2(\Delta t) = \lambda\gamma\pi_2(0)\Delta t + o(\Delta t) > 0$ hence we can assume that $\beta_2(0) > 0$. Similarly we can assume that $\pi_3(0) > 0$ and $\beta_3(0) > 0$.

2. Suppose that $\pi(0) = 0$. Hence $\beta(0) > 0$. Following the same method as the previous case we find that

$$\begin{aligned} \pi(\Delta t) &= \lambda\hat{\beta}(0)(1-\phi)\Delta t + o(\Delta t) > 0, \\ \beta(\Delta t) &= \beta(0) - (\beta(0)\lambda\gamma + \beta(0)\tau)\Delta t + o(\Delta t) > 0, \qquad \text{(for } \Delta t \text{ small enough),} \end{aligned}$$

and

$$\psi(\Delta t) = 1 - \lambda\hat{\beta}(0)(1-\phi)\Delta t + o(\Delta t) > 0, \qquad \text{(for } \Delta t \text{ small enough).}$$

Hence by choosing Δt small enough and starting at $t = \Delta t$ without loss of generality we can assume that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$ and as in the previous case we can also assume without loss of generality that $\pi_1(0) > 0$, $\beta_1(0) > 0$, $\pi_2(0) > 0$, $\beta_2(0) > 0$, $\pi_3(0) > 0$ and $\beta_3(0) > 0$.

3. Suppose that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$. This case is trivial and follows directly from the argument in Case 1.

4. Suppose that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) = 0$. This implies that $\pi(0) = 1$, hence

$$\psi(\Delta t) \geq \mu\Delta t + o(\Delta t) > 0, \qquad \text{(for } \Delta t \text{ small and positive).}$$

Thus it follows directly that without loss of generality we can assume that $\psi(0) > 0$ and the result follows by Case 1.

This completes the proof of Lemma 3.1.•

From Lemma 3.1 we have that there exists fixed ϵ where $1 > \epsilon > 0$ such that if Δt is small enough $\pi_i(\Delta t) \geq \epsilon\pi_i^\star$ and $\beta_i(\Delta t) \geq \epsilon\beta_i^\star$ for $i = 1, 2, 3$. As in the Simple Model suppose first that $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$. Then arguing as for the Simple Model we have that there exist T_1, T_2 and T_3 such that for $t \geq \max(T_1, T_2, T_3)$, $\pi_1 \geq \frac{1}{4}\epsilon\pi_1^\star$, $\pi_2 \geq \frac{1}{4}\epsilon\pi_2^\star$ and $\pi_3 \geq \frac{1}{4}\epsilon\pi_3^\star$. From equations (3.18)-(3.20) we have that

$$\begin{aligned}
\beta_{1,\infty} &\geq \frac{\lambda\gamma}{\lambda\gamma+\tau}\pi_{1,\infty} \geq \frac{1}{2}\epsilon\frac{\lambda\gamma\pi_1^\star}{\lambda\gamma+\tau} = \frac{1}{2}\epsilon\beta_1^\star, \\
\beta_{2,\infty} &\geq \frac{\lambda\gamma}{\lambda\gamma+\tau}\frac{\delta_1}{\mu+\delta_2}\pi_{1,\infty} \geq \frac{1}{2}\epsilon\frac{\lambda\gamma}{\lambda\gamma+\tau}\frac{\delta_1\pi_1^\star}{\mu+\delta_2} = \frac{1}{2}\epsilon\beta_2^\star, \\
\text{and} \quad \beta_{3,\infty} &\geq \frac{\lambda\gamma}{\lambda\gamma+\tau}\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\pi_{1,\infty} \geq \frac{1}{2}\epsilon\frac{\lambda\gamma}{\lambda\gamma+\tau}\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\pi^\star = \frac{1}{2}\epsilon\beta_3^\star.
\end{aligned}$$

Hence arguing again as in the Simple Model there exists T_4, T_5 and T_6 such that for $t \geq \max(T_4, T_5, T_6)$, $\beta_1 \geq \frac{1}{4}\epsilon\beta_1^\star$, $\beta_2 \geq \frac{1}{4}\epsilon\beta_2^\star$ and $\beta_3 \geq \frac{1}{4}\epsilon\beta_3^\star$. So for $t \geq \max(T_1, T_2, T_3, T_4, T_5, T_6)$ we have that $\pi_1 \geq \frac{1}{4}\epsilon\pi_1^\star$, $\pi_2 \geq \frac{1}{4}\epsilon\pi_2^\star$, $\pi_3 \geq \frac{1}{4}\epsilon\pi_3^\star$, $\beta_1 \geq \frac{1}{4}\epsilon\beta_1^\star$, $\beta_2 \geq \frac{1}{4}\epsilon\beta_2^\star$ and $\beta_3 \geq \frac{1}{4}\epsilon\beta_3^\star$, in other words eqn (3.12) holds true with ϵ replaced by $\epsilon/4$.

Now suppose that $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$ in which case there exists $\zeta \geq \Delta t$ such that $\pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star$. Let $t_0 = \inf\{\zeta \geq \Delta t, \pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star\}$ and $t_1 = \inf\{\zeta \geq t_0, \pi_1(\zeta) > \frac{1}{2}\epsilon\pi_1^\star\}$ where ϵ is fixed and positive. By the definition of t_0 we have that $\pi_1(t_0+\nu) < \frac{1}{2}\epsilon\pi_1^\star$ if ν is small and positive, hence $t_1 > t_0$. As in the Simple Model by continuity $\pi_1(t_0) = \pi_1(t_1) = \frac{1}{2}\epsilon\pi_1^\star$ and π_1 is less than $\frac{1}{2}\epsilon\pi_1^\star$ in (t_0, t_1) and greater than $\frac{1}{2}\epsilon\pi_1^\star$ just after t_1. We now show that if π_1 becomes small then all components become small.

By again exploiting the similarities between the Optimistic and Simple Models we can use Lemma 2.4 and Corollary 2.5 directly. From eqns (3.4) -(3.6) we have that if $\bar{T}_1$ and $\bar{T}_2$ are defined as for the Simple Model and Δ is small and positive then

$$\begin{aligned}
&\text{for } t_0 \in [t_0, t_1] && \frac{d\beta_1}{dt} \leq \frac{1}{2}\epsilon\pi_1^\star\lambda\gamma - (\lambda\gamma+\tau)\beta_1, \\
&\text{for } t_0 \in [t_0+\bar{T}_1, t_1] && \frac{d\beta_2}{dt} \leq \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star\lambda\gamma - (\lambda\gamma+\tau)\beta_2, \\
&\text{and for } t_0 \in [t_0+\bar{T}_1+\bar{T}_2, t_1] && \frac{d\beta_3}{dt} \leq \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star\lambda\gamma - (\lambda\gamma+\tau)\beta_3.
\end{aligned}$$

Hence we have that there exists a time $\bar{T}_3$ such that for all $t \geq t_0 + \bar{T}_3$, $0 < \beta_1 < (\frac{1}{2}+\Delta)\beta_1^\star\epsilon$. Similarly there exists a time $\bar{T}_4$ such that for all $t \geq t_0+\bar{T}_1+\bar{T}_4$, $0 < \beta_2 < (\frac{1}{2}+2\Delta)\beta_2^\star\epsilon$, and a time $\bar{T}_5$ such that for all $t \geq t_0+\bar{T}_1+\bar{T}_2+\bar{T}_5$, $0 < \beta_3 < (\frac{1}{2}+3\Delta)\beta_3^\star\epsilon$. In each case the value of $\bar{T}_i$ depends only on the Optimistic Model parameters, ϵ and Δ.

We have shown that if π_1 approaches zero then all components must also approach zero. We now show that π_1 cannot become arbitrarily small. We do this by showing that $t_1 - t_0$ can be bounded above by a fixed finite value and hence π_1 is not below $\frac{1}{2}\epsilon\pi_1^\star$ long enough to become arbitrarily close to zero. Now either π_1 remains below $\frac{1}{2}\epsilon\pi_1^\star$ long enough for all components to become small or π_1 increases up past $\frac{1}{2}\epsilon\pi_1^\star$ before all components become small. Hence we have that either

$$(i) \qquad t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5],$$

or $$(ii) \qquad t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5].$$

We wish to show that $t_1 < T$ where T is a fixed finite value dependent only on the model parameters, ϵ and Δ. If case (ii) is true then we are finished. Case (i) is where all components become small and it is this situation we now deal with. Since the disease-free equilibrium is unstable for $R_0 > 1$ we can use this to show that π_1 cannot stay small indefinitely.

Lemma 3.2 *If $\pi_1(t)$ drops to below $\frac{1}{2}\epsilon\pi_1^\star$ at time t_0 then $\pi_1(t)$ returns to $\frac{1}{2}\epsilon\pi_1^\star$ by at least time $t_1^+ = t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5, t_2 + \bar{T}_6]$ where $t_1^+ - t_0$ is finite and depends only on Δ, ϵ and the model parameters.*

Proof.

Suppose that ϵ_2 is real and positive and $1 \geq \epsilon_2 \geq 0$ and consider the matrix $\mathbf{J}(\epsilon_2) =$

$$\begin{bmatrix}
-(\mu+\delta_1) & 0 & 0 & \lambda\alpha_1(1-\phi)(1-\epsilon_2) & \lambda\alpha_2(1-\phi)(1-\epsilon_2) & \lambda\alpha_3(1-\phi)(1-\epsilon_2) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\
\lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 & 0 \\
0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 \\
0 & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau)
\end{bmatrix}.$$

When $\epsilon_2 = 0$, $\mathbf{J}(0) = \mathbf{J}$, the linearised stability matrix about the disease-free equilibrium as used in the proof of Theorem 3.3. Suppose that the eigenvalues of $\mathbf{J}(\epsilon_2)$ are $w_1(\epsilon_2)$, $w_2(\epsilon_3)$, $w_3(\epsilon_2)$, $w_4(\epsilon_2)$, $w_5(\epsilon_2)$ and $w_6(\epsilon_2)$. Note that the only negative entries in $\mathbf{J}(\epsilon_2)$ are the ones on the leading diagonal. Arguing as in the proof of Lemma 2.5 without loss of generality we may assume that $w_1(\epsilon_2)$ is real and the other eigenvalues have strictly smaller real parts. In particular this is true when $\epsilon_2 = 0$. Moreover from Corollary 2.7 we have that the roots of the characteristic equation of $\mathbf{J}(\epsilon_2)$ are continuous functions of ϵ_2, hence $w_1(\epsilon_2) \to w_1(0)$ as $\epsilon_2 \to 0$. From the proof of Theorem 3.3

we know that $w_1(0)$ is strictly positive when $R_0 > 1$. Therefore by choosing ϵ_2 small enough we can ensure that $w_1(\epsilon_2) > 0$. Without loss of generality we can assume that $1 > \epsilon_2 > 0$. We can choose ϵ small enough such that

$$\frac{1}{2}\epsilon\pi_1^\star + \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star + \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star \;<\; \epsilon_2.$$

Hence for $t_1 > t \geq t_0 + \bar{T}_1 + \bar{T}_2$ we have that $\pi_1 + \pi_2 + \pi_3 < \epsilon_2$ using Lemma 2.4 and Corollary 2.5. Let $t_2 = \inf\{\zeta : \text{ for } t_1 > t \geq t_0 + \zeta, \pi(t) < \epsilon_2\}$, hence if $t_2 > 0$, $\pi(t_0 + t_2) = \epsilon_2$ and $t_0 + t_2$ is the last time before t_1 that $\pi(t) \geq \epsilon_2$, and note that $t_2 \leq \bar{T}_1 + \bar{T}_2$. If $t_1 < t_0 + \bar{T}_1 + \bar{T}_2$ then we have the desired result. Now we consider the case where $t_1 \geq t \geq t_0 + \bar{T}_1 + \bar{T}_2$. We have that

$$\frac{d\mathbf{x}}{dt} \;\geq\; \mathbf{J}(\epsilon_2)\mathbf{x},$$

where $\mathbf{x} = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$. From Lemma 2.1 in Nold (1980), $\mathbf{J}(\epsilon_2)$ has a strictly positive left eigenvector, $\mathbf{e} = (e_1, e_2, e_3, e_4, e_5, e_6)$ corresponding to the Perron Eigenvalue $w_1(\epsilon_2)$. Hence

$$\mathbf{e}\,\frac{d\mathbf{x}}{dt} \;\geq\; \mathbf{e}\,\mathbf{J}(\epsilon_2)\mathbf{x} \;=\; w_1(\epsilon_2)\,\mathbf{e}.\mathbf{x}.$$

Thus

$$\begin{aligned}
\mathbf{e}.\mathbf{x}(\mathbf{t}) \;&\geq\; \mathbf{e}.\mathbf{x}(t_0+t_2)\exp[w_1(\epsilon_2)(t-t_0-t_2)], && \text{(integrating over}[t_0+t_2, t]),\\
&\geq\; (e_1\pi_1 + e_2\pi_2 + e_3\pi_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)],\\
&\geq\; \pi(t_0+t_2)\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)],\\
&\begin{cases} =\; \epsilon_2\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)], & \text{if } t_2 > 0,\\ \geq\; \frac{1}{2}\epsilon\pi_1^\star\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)], & \text{if } t_2 = 0. \end{cases}
\end{aligned}$$

Therefore after a time $t_0 + t_2 + \bar{T}_6$ we have that

$$\mathbf{e}.\mathbf{x}(t) \;>\; \mathbf{e}.\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\beta_1^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\beta_2^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon\beta_3^\star\right),$$

where $\bar{T}_6$ depends only on $\epsilon_2, \epsilon, \Delta$ and the model parameters.

We know that provided that $t_0 \leq t \leq t_1$ then after a time $t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1+\bar{T}_4, \bar{T}_1+\bar{T}_2+\bar{T}_5]$, we have that $\pi_1 \leq \frac{1}{2}\epsilon\pi_1^\star$, $\pi_2 \leq (\frac{1}{2}+\Delta)\epsilon\pi_2^\star$, $\pi_3 \leq (\frac{1}{2}+2\Delta)\epsilon\pi_3^\star$, $\beta_1 \leq (\frac{1}{2}+\Delta)\epsilon\beta_1^\star$, $\beta_2 \leq (\frac{1}{2}+2\Delta)\epsilon\beta_2^\star$ and $\beta_3 \leq (\frac{1}{2}+3\Delta)\epsilon\beta_3^\star$. Hence

$$\mathbf{e}.\mathbf{x}(t) \;\leq\; \mathbf{e}.\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\beta_1^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\beta_2^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon\beta_3^\star\right).$$

However if $t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5, t_2 + \bar{T}_6]$ we have a contradiction. Hence $t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5, t_2 + \bar{T}_6]$. This concludes the proof of Lemma 3.2.•

As in the case of the Simple Model we have shown that the first time π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$ it must return back to this level by a duration of at most T time units later, and as before this extends to any time that π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$. Hence for the Optimistic Model we also have that if π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$ at $\tilde{t}_0$ then for $t \in [\tilde{t}_0, \tilde{t}_0 + T]$,

$$\begin{aligned} \frac{d\pi_1}{dt} &\geq -(\mu + \delta_1)\pi_1, \\ \pi_1 &\geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu + \delta_1)(t - \tilde{t}_0)], \\ &\geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu + \delta_1)T], \end{aligned}$$

where T is a fixed duration dependent only on $\epsilon, \epsilon_2, \Delta$ and the model parameters. Since $\frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu + \delta_1)T]$ is strictly positive we have that $\pi_{1,\infty} > 0$. Hence (3.12) is true and we have that there exists $\epsilon > 0$ and $\eta > 0$ such that for all $t \geq \eta$, $\pi_i(t) \geq \epsilon$ and $\beta_i(t) \geq \epsilon$ for $i = 1, 2, 3$. This completes the proof of Theorem 3.4.•

3.7.1 Local Stability of Endemic Equilibrium

Showing local stability of the endemic equilibrium in the Optimistic Model would require examining the Routh-Hurwitz conditions for a sixth order polynomial. This is very complicated and we feel that while theoretically possible the work required outweighs the merits of a local stability result. We instead examine local stability in a model which is closely related to the Optimistic Model and has the same endemic equilibrium solution but which has only three rather than six dimensions.

Evidence suggests that the timescale on which addicts inject is of the order of days, whereas that of the other epidemiological and demographic processes is measured in years, and is a lot slower. We commented that this appears to be true in the Simple Model in the previous chapter. For example by examining Figures 2.2-2.4 we can see that for each of the initial conditions shown $\beta(t)$ very quickly settles down to an almost parallel trajectory to $\pi(t)$. By examining the needle equations (3.4)-(3.6) in our Optimistic Model it is obvious that if the prevalence of disease is constant among the addict population (at say values of π_i for $i = 1, 2, 3$) then the prevalence of disease in each group of infectious needles will tend to $\pi_i/(1 + \hat{\tau})$ for $i = 1, 2, 3$ for state one, two and three infectious needles respectively. Now whilst it is not true that the π_i

values are fixed, it is true that the β_i's will respond very rapidly to slowly varying π_i-values, so we can approximate the dynamic relationship between $\pi_i(t)$ and $\beta_i(t)$ as $\beta_i(t) = \pi_i(t)/(1+\hat{\tau})$ for $i = 1,2,3$. A similar argument is used by Kaplan and O'Keefe (1993) to calculate the reduction in HIV incidence due to the introduction of a needle exchange in their model, and by Kaplan (1994, 1995) to justify assuming that HIV prevalence amongst addicts is constant in a model examining only the dynamics of HIV amongst needles. We later demonstrate using simulations that this approximation for the prevalence of disease in needles is extremely good. It appears that for any initial conditions the prevalence of disease in needles very quickly settles down to the approximate relationship $\beta_i(t) = \pi_i(t)/(1+\hat{\tau})$.

By assuming that $\beta_i(t) = \pi_i(t)/(1+\hat{\tau})$ for $i = 1,2,3$, we can model the prevalence of disease among addicts using only three variables, $\pi_1(t), \pi_2(t)$ and $\pi_3(t)$. This "addict only" model can be represented by the following system of equations:

$$\frac{d\pi_1}{dt} = (1-\pi)\lambda(1-\phi)\left(\frac{\pi_1\alpha_1 + \pi_2\alpha_2 + \pi_3\alpha_3}{1+\hat{\tau}}\right) - (\mu+\delta_1)\pi_1, \quad (3.21)$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2, \quad (3.22)$$

and

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3. \quad (3.23)$$

An important feature of eqns (3.21)-(3.23) is that by construction it has the same equilibrium solutions as in the full Optimistic Model.

The Jacobian at $(\pi^\star, \beta^\star)$ for eqns (3.21)-(3.23) is $\mathbf{J} = (j_{ij})$ where:

$$j_{11} = \frac{\lambda(1-\phi)}{1+\hat{\tau}}\Big[\alpha_1(1-2\pi^\star\eta_1) - \pi^\star\eta_2(\alpha_1+\alpha_2) - \pi^\star\eta_3(\alpha_1+\alpha_3)\Big] - (\mu+\delta_1); \quad (3.24)$$

$$j_{12} = \frac{\lambda(1-\phi)}{1+\hat{\tau}}\Big[\alpha_2(1-2\pi^\star\eta_2) - \pi^\star\eta_1(\alpha_1+\alpha_2) - \pi^\star\eta_3(\alpha_2+\alpha_3)\Big]; \quad (3.25)$$

$$j_{13} = \frac{\lambda(1-\phi)}{1+\hat{\tau}}\Big[\alpha_3(1-2\pi^\star\eta_3) - \pi^\star\eta_1(\alpha_1+\alpha_3) - \pi^\star\eta_2(\alpha_2+\alpha_3)\Big]; \quad (3.26)$$

$j_{21} = \delta_1$; $j_{22} = -(\mu+\delta_2)$; $j_{23} = 0$; $j_{31} = 0$; $j_{32} = \delta_2$; and $j_{33} = -(\mu+\delta_3)$. Recall that for the Optimistic Model $\pi^\star = 1 - (1/R_0)$ and $\pi_i^\star = \eta_i\pi^\star$ for $i = 1,2,3$. If the characteristic equation is denoted by $\omega^3 + a_1\omega^2 + a_2\omega + a_3 = 0$ then it is easy to show that

$$a_1 = (\mu+\delta_2+\mu+\delta_3) - j_{11}, \quad (3.27)$$

$$a_2 = (\mu+\delta_2)(\mu+\delta_3) - \delta_1 j_{12} - (\mu+\delta_2+\mu+\delta_3)j_{11}, \quad (3.28)$$

and

$$a_3 = -(\mu+\delta_2)(\mu+\delta_3)j_{11} - \delta_1(\mu+\delta_3)j_{12} - \delta_1\delta_2 j_{13}. \tag{3.29}$$

By substituting $1-(1/R_0)$ for $\pi^\star$ and after some simplification we find that

$$a_1 = (\mu+\delta_2+\mu+\delta_3) + \frac{\mu+\delta_1}{L}(R_0-1) + \frac{\mu+\delta_1}{F}\left(\frac{\alpha_2\delta_1}{\mu+\delta_2} + \frac{\alpha_3\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right),$$

$$\begin{aligned} a_2 = {} & (\mu+\delta_2)(\mu+\delta_3) + \frac{\mu+\delta_1}{L}(\delta_1+\mu+\delta_2+\mu+\delta_3)(R_0-1) \\ & + \frac{\mu+\delta_1}{F}\left(\frac{\alpha_2\delta_1(\mu+\delta_3)}{\mu+\delta_2} + \frac{\alpha_3\delta_1\delta_2}{\mu+\delta_2} + \frac{\alpha_3\delta_1\delta_2}{\mu+\delta_3}\right), \end{aligned}$$

and

$$a_3 = \frac{\mu+\delta_1}{L}(\delta_1\delta_2+\delta_1(\mu+\delta_3)+(\mu+\delta_2)(\mu+\delta_3))(R_0-1),$$

where

$$F = \alpha_1 + \frac{\delta_1\alpha_2}{\mu+\delta_2} + \frac{\delta_1\delta_2\alpha_3}{(\mu+\delta_2)(\mu+\delta_3)}.$$

It is clear that if $R_0 > 1$ then $a_1 > 0$, $a_2 > 0$ and $a_3 > 0$, moreover it is also apparent that there are sufficient terms in a_1a_2 such that $a_1a_2 > a_3$. Hence the Routh-Hurwitz conditions for a cubic are satisfied and the endemic equilibrium in our "addict only" model is locally stable.

It is possible to make this approximation argument more rigorous by showing that if $\lambda\gamma$ is large compared with the other parameters of the model apart from τ, (including $\lambda\alpha_1$, $\lambda\alpha_2$ and $\lambda\alpha_3$) then three of the roots of the characteristic equation of the Jacobian of the full model at the endemic equilibrium are close to $-(\lambda\gamma+\tau)$ and the other three are close to the roots of the characteristic equation of the Jacobian of the "addict-only" model at the endemic equilibrium. See Appendix C for details.

3.7.2 Sufficient Conditions for Global Stability of $(\pi^\star, \beta^\star)$

We now use a similar method to that discussed in Section 2.5.1 to derive sufficient conditions for the global stability of the endemic equilibrium in the Optimistic Model. We again consider a translated form of our original coordinate system, $(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$, where the origin of the new system corresponds to the endemic equilibrium, $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star)$, in the original form. This translation gives the following set of model equations:

$$\frac{d\tilde{\pi}_1}{dt} = (1-\pi^\star)\lambda(1-\phi)(\tilde{\beta}_1\alpha_1+\tilde{\beta}_2\alpha_2+\tilde{\beta}_3\alpha_3) - \tilde{\pi}\lambda(1-\phi)(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)$$

$$-(\mu+\delta_1)\tilde{\pi}_1, \tag{3.30}$$

$$\frac{d\tilde{\pi}_2}{dt} = \delta_1\tilde{\pi}_1 - (\mu+\delta_2)\tilde{\pi}_2, \tag{3.31}$$

$$\frac{d\tilde{\pi}_3}{dt} = \delta_2\tilde{\pi}_2 - (\mu+\delta_3)\tilde{\pi}_3, \tag{3.32}$$

$$\frac{d\tilde{\beta}_1}{dt} = \lambda\gamma(\tilde{\pi}_1 - \tilde{\beta}_1) - \tilde{\beta}_1\tau, \tag{3.33}$$

$$\frac{d\tilde{\beta}_2}{dt} = \lambda\gamma(\tilde{\pi}_2 - \tilde{\beta}_2) - \tilde{\beta}_2\tau, \tag{3.34}$$

$$\text{and } \frac{d\tilde{\beta}_3}{dt} = \lambda\gamma(\tilde{\pi}_3 - \tilde{\beta}_3) - \tilde{\beta}_3\tau, \tag{3.35}$$

where $\tilde{\beta}_i = \beta_i - \beta_i^\star$, $\tilde{\pi}_i = \pi_i - \pi_i^\star$ for $i = 1,2,3$, and $\tilde{\pi} = \tilde{\pi}_1 + \tilde{\pi}_2 + \tilde{\pi}_3$. We wish to show that $\tilde{\pi}_i \to 0$ and $\tilde{\beta}_i \to 0$ for $i = 1,2,3$, as $t \to \infty$. This is equivalent to showing that $\pi_i \to \pi_i^\star$ and $\beta_i \to \beta_i^\star$ for $i = 1,2,3$, as $t \to \infty$. The system defined by eqns (3.30)-(3.35) can be represented in matrix form as $d\tilde{\mathbf{x}}/dt = \mathbf{V}(\mathbf{x})\tilde{\mathbf{x}}$, where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$, $\tilde{\mathbf{x}}^T = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$ and $\mathbf{V}(\mathbf{x}) =$

$$\begin{bmatrix}
-(\mu+\delta_1)-\lambda\hat{\beta}(1-\phi) & -\lambda\hat{\beta}(1-\phi) & -\lambda\hat{\beta}(1-\phi) & (1-\pi^*)\lambda\alpha_1(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\
\lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 & 0 \\
0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 \\
0 & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau)
\end{bmatrix},$$

where $\hat{\beta} = \beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$. When $\mathbf{x} = \mathbf{0}$ we have that $\mathbf{V}(0) =$

$$\begin{bmatrix}
-(\mu+\delta_1) & 0 & 0 & (1-\pi^*)\lambda\alpha_1(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\
\lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 & 0 \\
0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 \\
0 & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau)
\end{bmatrix},$$

and note that the only strictly negative entries are on the leading diagonal.

We are also interested in the additional coordinate system of $(\tilde{\pi}, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$. As in the Simple Model the equation for $d\tilde{\pi}/dt$ is easily derived by adding eqns (3.30)-(3.32). This new system can be represented in matrix form as $d\tilde{\mathbf{y}}/dt = \mathbf{W}(\mathbf{y})\tilde{\mathbf{y}}$, where

$\mathbf{y}^T = (\pi, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$, $\tilde{\mathbf{y}}^T = (\tilde{\pi}, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$ and $\mathbf{W}(\mathbf{y}) =$

$$\begin{bmatrix} -\mu-\lambda\hat{\beta}(1-\phi) & 0 & -\delta_3 & (1-\pi^*)\lambda\alpha_1(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\ \delta_1 & -(\mu+\delta_2+\delta_1) & -\delta_1 & 0 & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\ \lambda\gamma & -\lambda\gamma & -\lambda\gamma & -(\lambda\gamma+\tau) & 0 & 0 \\ 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) & 0 \\ 0 & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau) \end{bmatrix}.$$

In this coordinate system the variable $\hat{\beta}$ only appears on the leading diagonal of $\mathbf{W}(\mathbf{y})$.

We can write $\tilde{\mathbf{y}} = \mathbf{J}\tilde{\mathbf{x}}$ where

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

hence using a similar method as in the proof of Lemma 2.6 we have that $\mathbf{V}(\mathbf{x})$ and $\mathbf{W}(\mathbf{y})$ share the same eigenvalues.

Using Theorem 3.4 we can replace the variable $\hat{\beta} = \beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$ in $\mathbf{W}(\mathbf{y})$ with a constant lower bound, $\bar{\epsilon}$. Hence we have that for $t \geq \eta$, $\mathbf{W}(\mathbf{y}) \leq \mathbf{W}^+ = \mathbf{W}(0) - \bar{\epsilon}\mathbf{E}$ where

$$\mathbf{E} = \begin{bmatrix} \lambda(1-\phi) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now following a similar argument as in the Simple Model, we now show that $\mathbf{W}(0) - \bar{\epsilon}\mathbf{E}$ is Lyapunov stable. As in the proof of Lemma 2.6 the eigenvalues of $\mathbf{W}(0)$ are the same as those for $\mathbf{V}(0)$. As previously if M is large enough $\mathbf{V}(0) + M\mathbf{I}$ is an irreducible matrix with non-negative elements and has a unique strictly positive eigenvector $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star)$. From Lemma 2.1 in Nold (1980), we find that the eigenvalue corresponding to the eigenvector $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star)$ is a simple eigenvalue and it is also the spectral radius of $\mathbf{V}(0) + M\mathbf{I}$. Thus all eigenvalues of $\mathbf{V}(0)$ and hence

$\mathbf{W(0)}$ lie in a circle centered on $(-M,0)$ with radius M, and zero is a simple eigenvalue of $\mathbf{W(0)}$.

The characteristic equation of $\mathbf{W}^+ = \mathbf{W(0)} - \bar{\varepsilon}\mathbf{E}$, is of the form

$$\omega^6 + a_1(\bar{\varepsilon})\omega^5 + a_2(\bar{\varepsilon})\omega^4 + a_3(\bar{\varepsilon})\omega^3 + a_4(\bar{\varepsilon})\omega^2 + a_5(\bar{\varepsilon})\omega + a_6(\bar{\varepsilon}) = 0, \tag{3.36}$$

where $a_i(\bar{\varepsilon})$, for $i = 1,\dots 6$, are continuous functions of $\bar{\varepsilon}$. We find that

$$\begin{aligned} a_6(\bar{\varepsilon}) \;=\; & \Big(\mu + \lambda\bar{\varepsilon}(1-\phi)\Big)\Big[(\mu+\delta_1+\delta_2)(\mu+\delta_3)+\delta_1\delta_2\Big](\lambda\gamma+\tau)^3 + \delta_1\delta_2\delta_3(\lambda\gamma+\tau)^3 \\ & -\alpha_1\lambda(1-\phi)(1-\pi^*)\lambda\gamma(\lambda\gamma+\tau)^2(\mu+\delta_3)(\mu+\delta_2) \\ & -\alpha_2\lambda(1-\phi)(1-\pi^*)\lambda\gamma(\lambda\gamma+\tau)^2(\mu+\delta_3)\delta_1 \\ & -\alpha_3\lambda(1-\phi)(1-\pi^*)\lambda\gamma(\lambda\gamma+\tau)^2\delta_1\delta_2. \end{aligned}$$

When $\bar{\varepsilon} = 0$ we have that $\mathbf{W}^+ = \mathbf{W(0)}$ and we know that zero is an eigenvalue of $\mathbf{W(0)}$. When $\bar{\varepsilon} = 0$ five eigenvalues have strictly negative real parts and the sixth eigenvalue is real and lies at the origin. Now consider the case where $\bar{\varepsilon} > 0$, as in the Simple Model we have that the eigenvalues are continuous in $\bar{\varepsilon}$ in a neighbourhood about the origin. By continuity ω_i for $i = 1,\dots,5$ will still have strictly negative real parts. Now consider what happens to ω_6 as $\bar{\varepsilon}$ increases from zero to a small positive value. If ω_6 either stays on the imaginary axis or moves to the right and gains a positive real part then we have that $a_6 = \omega_1\omega_2\omega_3\omega_4\omega_5\omega_6 \leq 0$ however this is impossible since we know that a_6 is strictly increasing in $\bar{\varepsilon}$ and that $a_6(0) = 0$. Hence for $\bar{\varepsilon} > 0$ small and positive all eigenvalues of $\mathbf{W}^+$ have strictly negative real parts. Therefore $\mathbf{W}^+$ is Lyapunov stable.

The forms of $\mathbf{W(y)}$ and $\mathbf{W}^+$ are similar to those discussed in Section 2.5.1 and therefore we again require the stronger condition that $\mathbf{W}^+$ is Volterra-Lyapunov stable (also known as diagonal stability) to guarantee that if disease is initially present then the system will approach the endemic equilibrium when $R_0 > 1$.

Recall the definition of $\mathcal{M}(\mathbf{A})$ on page 68. By considering the matrix $\mathbf{S} = -\mathbf{W}^{+T}$ we find that $\mathcal{M}(\mathbf{S}) =$

$$\begin{bmatrix} \mu+\lambda\epsilon(1-\phi) & -\delta_1 & 0 & -\lambda\gamma & 0 & 0 \\ 0 & \mu+\delta_2+\delta_1 & -\delta_2 & -\lambda\gamma & -\lambda\gamma & 0 \\ -\delta_3 & -\delta_1 & \mu+\delta_3 & -\lambda\gamma & 0 & -\lambda\gamma \\ -(1-\pi^*)\lambda\alpha_1(1-\phi) & 0 & 0 & \lambda\gamma+\tau & 0 & 0 \\ -(1-\pi^*)\lambda\alpha_2(1-\phi) & 0 & 0 & 0 & \lambda\gamma+\tau & 0 \\ -(1-\pi^*)\lambda\alpha_3(1-\phi) & 0 & 0 & 0 & 0 & \lambda\gamma+\tau \end{bmatrix}.$$

A sufficient condition for Volterra-Lyapunov stability is that all the principal minors of $\mathcal{M}(\mathbf{S})$ are strictly positive. By examining each of the individual principal minors it is straightforward to verify that $(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2>0$ and $\det(\mathcal{M}(\mathbf{S}))>0$ are necessary and sufficient conditions for the positivity of all principal minors. We find that $\det(\mathcal{M}(\mathbf{S}))=$

$$-\delta_1(1-\pi^\star)\lambda\alpha_3(1-\phi)(\lambda\gamma+\tau)^2\delta_2\lambda\gamma-\delta_1(\lambda\gamma+\tau)^2(1-\pi^\star)\lambda\alpha_2(1-\phi)\lambda\gamma(\mu+\delta_3)$$

$$-(\lambda\gamma+\tau)^2(1-\pi^\star)\lambda\alpha_1(1-\phi)\delta_1\lambda\gamma(\delta_2+\mu+\delta_3)-\delta_1\delta_2\delta_3(\lambda\gamma+\tau)^3$$

$$-\lambda\gamma(\lambda\gamma+\tau)^2(1-\pi^\star)\lambda\alpha_1(1-\phi)\Big[(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2\Big]$$

$$+(\lambda\gamma+\tau)^3(\mu+\lambda\bar{\varepsilon}(1-\phi))\Big[(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2\Big]. \qquad (3.37)$$

Therefore if these conditions are satisfied and $R_0>1$ then provided that disease is initially present $\pi\to\pi^\star$ and $\beta\to\beta^\star$ as $t\to\infty$. As in the Simple Model the condition $\det(\mathcal{M}(\mathbf{S}))>0$ is a genuine condition in that it is true for some parts of the parameter space but not others. For example if $\mu\gg\delta_1,\delta_2,\delta_3,\lambda\alpha_1$ then it will be true. However if $\bar{\varepsilon}$ is near zero and $\alpha_1>0$, $\alpha_2=\alpha_3=0$ then this condition together with $(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2>0$ implies that

$$\mu \;>\; (1-\pi^\star)\frac{\lambda\alpha_1(1-\phi)}{1+\hat{\tau}},$$

whereas the equilibrium equations imply that

$$(\mu+\delta_1)\pi_1^\star \;=\; (1-\pi^\star)\frac{\lambda\alpha_1(1-\phi)\pi_1^\star}{1+\hat{\tau}},$$

so

$$\mu \;<\; (1-\pi^\star)\frac{\lambda\alpha_1(1-\phi)}{1+\hat{\tau}}.$$

We have demonstrated a number of properties of the Optimistic Model using analytical results. We now move on to using numerical simulation to examine the dynamic behaviour of our model. We are also particularly interested in whether the prevalence of disease does indeed tend to the unique endemic equilibrium if disease is initially present and $R_0>1$. We have been unable to show analytically that $R_0>1$ (with disease initially present) is strong enough on its own to guarantee this behaviour but our simulation results suggest that this will be the case.

3.8 Simulation Study of the Optimistic Model

In order to validate the theoretical results of the previous sections we now demonstrate the behaviour of the Optimistic Model graphically using numerical simulations. We once again adopt a similar approach as for the Simple Model, we use the same parameter estimates (where appropriate) in our simulations and create the same kind of plots. It is important to note that the following simulations are *not* suitable for comparison with the Simple Model. Comparing and contrasting the characteristics of these models is obviously important but requires a method of calibrating the models in order to provide meaningful comparisons. We discuss this in Chapter 7.

The Optimistic Model incorporates three types of infectious needles and therefore we need to estimate the probability of transmission of HIV in a single shared injection from each type of infectious needle. We have denoted the HIV transmission probabilities as α_1, α_2 and α_3 from a single injection with a needle in state one, state two and state three infectivity respectively. The parameter α was the equivalent parameter in the Simple Model. We have assumed that addicts in each infectious stage have different concentrations of HIV in their blood. Using data on viral antigen recovery and epidemiological data from transfusion recipients Peterson et al. (1990) estimate that the relative HIV viral load of addicts exists in the ratio 5:1:3 for Acute Infection:Asymptomatic:Pre-AIDS Symptoms. Hence we assume that a needle used by an addict in stage one (Acute Infection) will be left fives times more infectious than if the addict were in stage two (Asymptomatic). Similarly we assume that a needle used by an addict in stage three (Pre-AIDS Symptoms) will be three times more infectious than if the addict were in stage two. We therefore assume that $\alpha_1 = \zeta_1\alpha_2$, $\alpha_3 = \zeta_3\alpha_2$ and that α_1:α_2:α_3 exist in the ratio 5:1:3. Hence we only need to estimate the baseline transmission probability, α_2. This parameter can be estimated using a similar method to that used to estimate α (see Appendix B for details). Table 3.1 contains a summary of the parameter estimates we shall use to simulate the Optimistic Model and other three stage infectivity models which we discuss in later chapters.

As in the Simple Model we wish to demonstrate two key properties of the Optimistic Model. If the parameter estimates are such that $R_0 > 1$ then provided that the disease is present in at least one addict or one needle then the disease spreads among the population until a steady state is reached where a fraction $\pi^\star$ of all addicts and a fraction $\beta^\star$ of all needles are infected with the disease where $\pi^\star$ and $\beta^\star$ are given by eqn

Table 3.1: Summary of Parameter Estimates for Three Stage Models

Parameter	Estimate
λ	246.22 per year
γ	0.90797
α_1	0.01412 per shared injection
α_2	0.00282 per shared injection
α_3	0.00847 per shared injection
μ	0.1333 per year
ϕ	0.64
τ	15.531 per year
δ_1	4.6154 per year
δ_2	0.2281 per year
δ_3	0.1920 per year

(3.11). We also wish to demonstrate that if the parameter estimates give rise to a value of $R_0 \leq 1$ then the disease will die out in both addicts and needles. We use two sets of parameters to illustrate these properties. The first set of parameters uses the estimates in Table 3.1 and gives a value for R_0 of 2.200. The second set of parameters is the same as the first with the exception of ϕ (the probability that an addict successfully cleans a needle prior to use) which is now 0.853, this gives a value for R_0 of 0.901.

Parameter Set One - $R_0 = 2.200$

We now simulate the Optimistic Model using the first set of parameter estimates where $R_0 = 2.2\dot{0}0$. Figure 3.1 shows the Optimistic Model simulated over seventy years. At time zero we have assumed that one percent of the total population of addicts are in stage one infectivity, at this time no other addicts or needles are infectious. The figure shows the progress of each type of infectious addict and each type of infectious needle over time. It is clear that the fraction of infected addicts in each stage eventually reaches a steady state as does the fraction of infected needles in each stage. We can also observe that the fraction of addicts and needles in the same infectious stage behave similarly. We know from the previous analytical work that if needles are never exchanged (which corresponds to $\tau = 0$) then the fraction of infected addicts and needles in the same

Figure 3.1: Optimistic Model when $R_0 > 1$

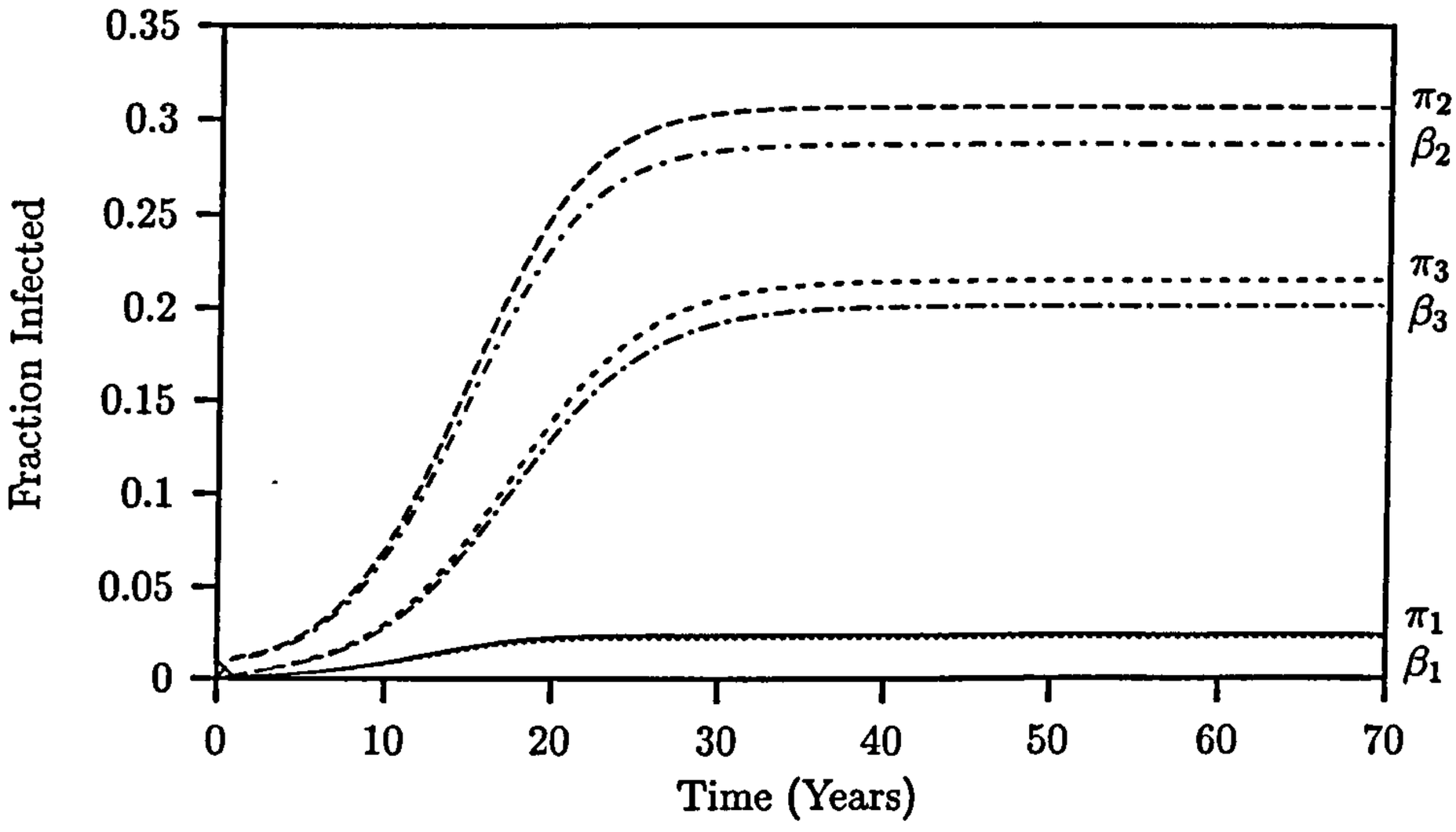


Figure 3.2: Optimistic Model when $R_0 > 1$ (Total Prevalence)

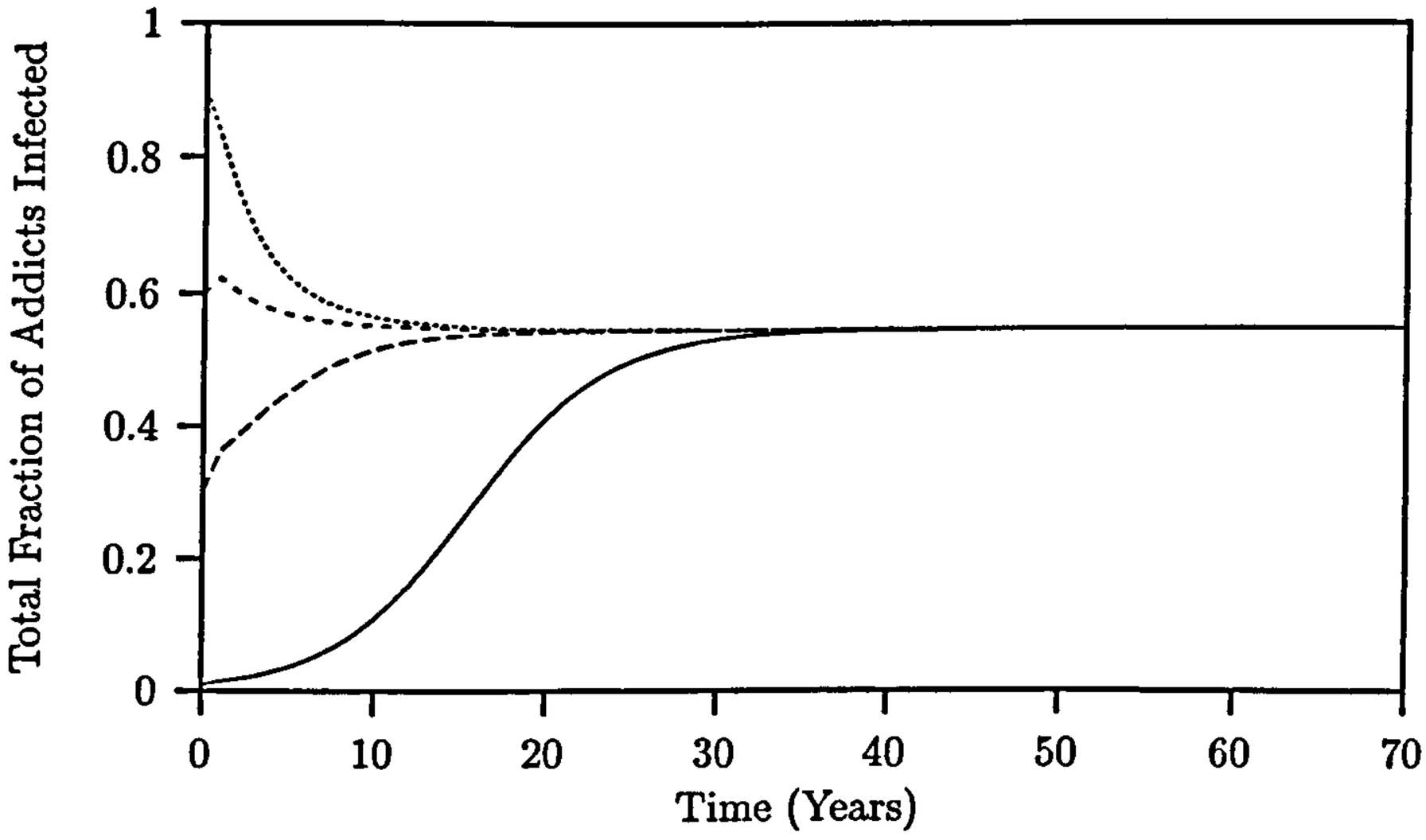
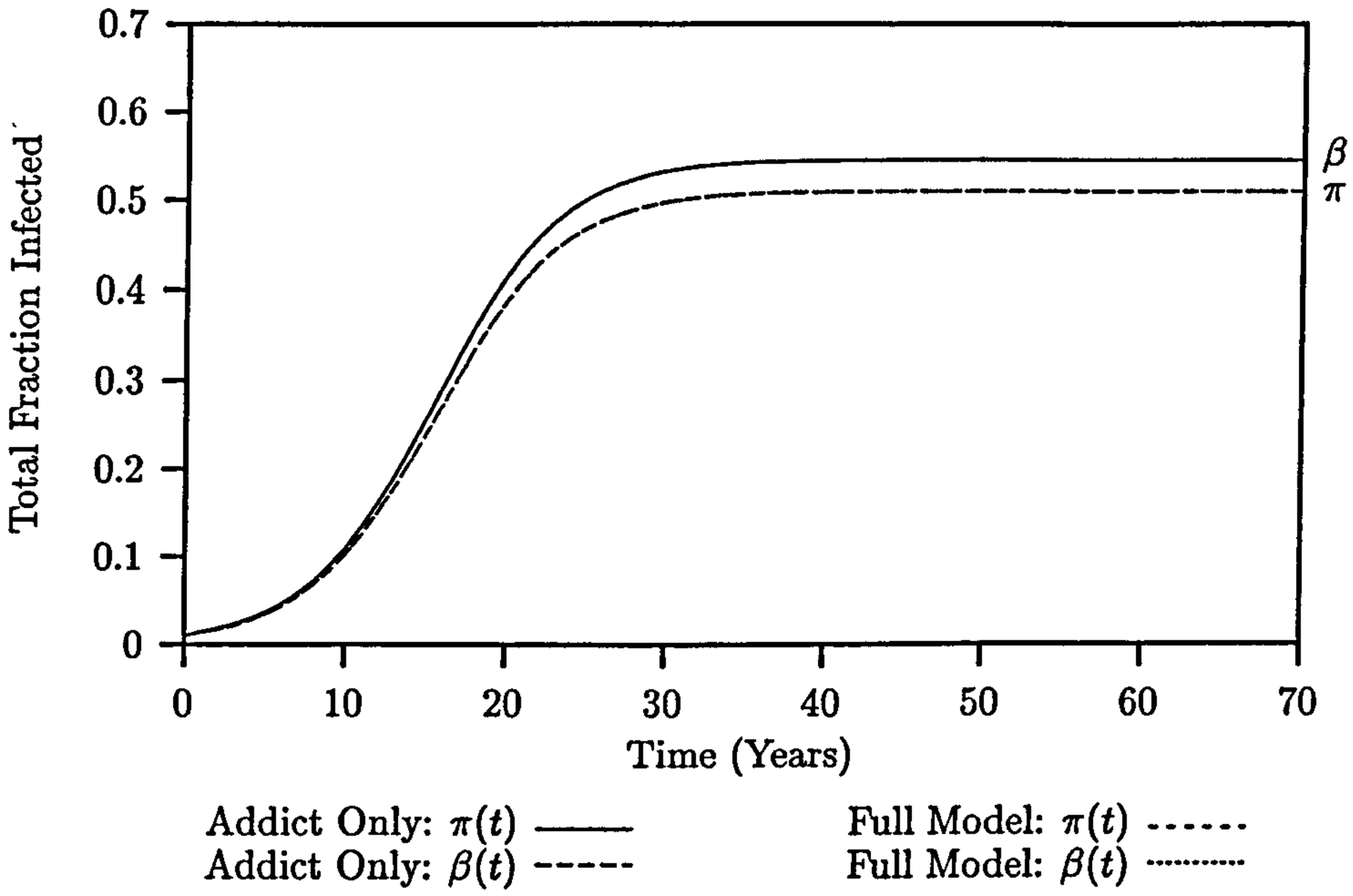


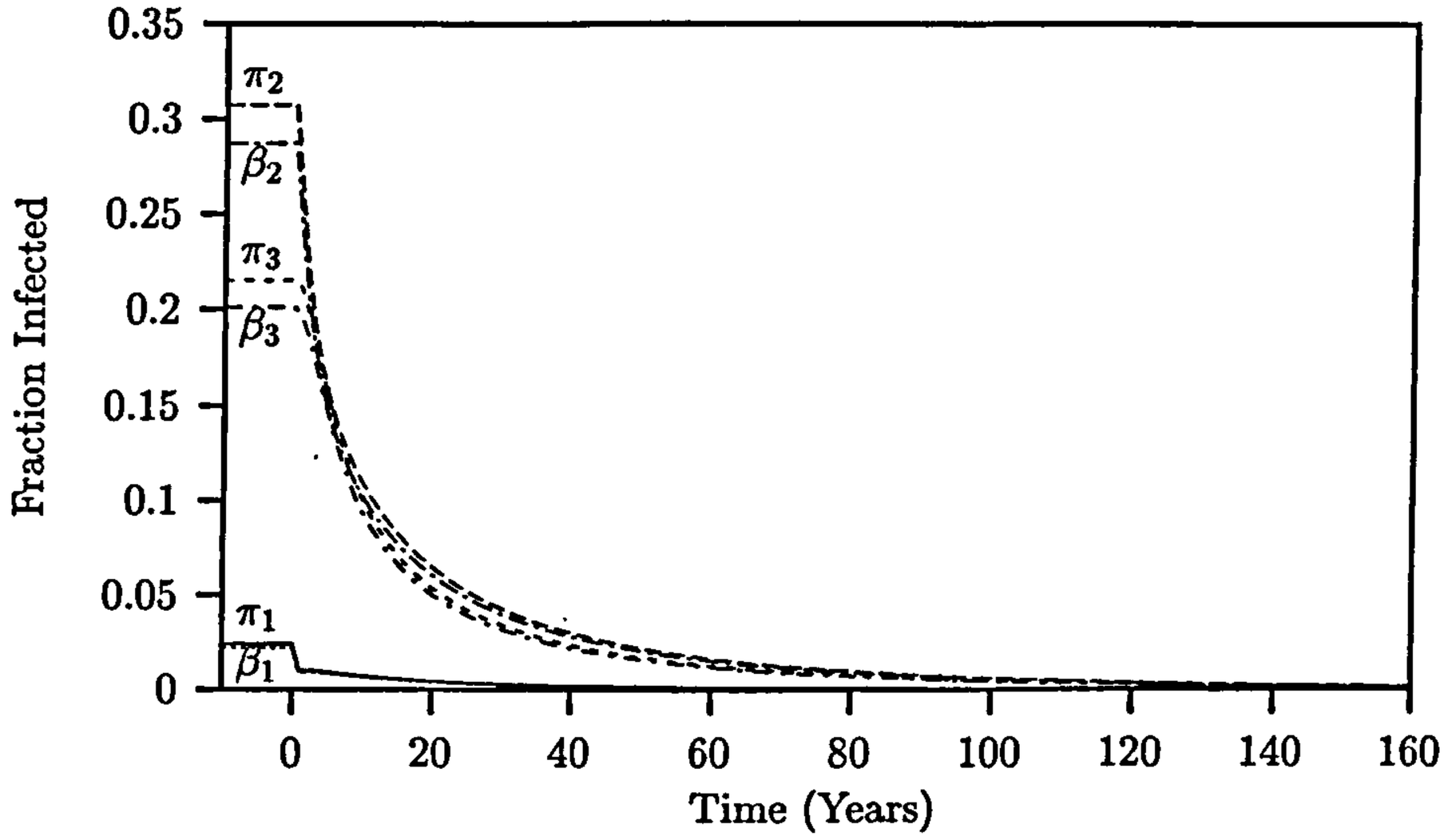
Figure 3.3: Optimistic (Addict only) Model when $R_0 > 1$



infectious stage are equal at equilibrium. The steady state values in these simulations are $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star) = (0.024, 0.307, 0.215, 0.022, 0.287, 0.201)$, which correspond to $\pi^\star = 0.545$ and $\beta^\star = 0.510$. Figure 3.2 illustrates the behaviour of the total prevalence of disease in addicts using same parameter estimates as in Figure 3.1 but for a variety of different initial conditions. In these trajectories the initial proportion of stage one infectious addicts takes the value indicated and no other addicts or needles are initially infected. A number of other simulations were carried out using various initial conditions and parameter estimates (where $R_0 > 1$). These simulations suggest that the Optimistic Model has a globally stable endemic equilibrium when $R_0 > 1$ and disease is initially present.

Before we illustrate the behaviour of the Optimistic Model for $R_0 < 1$ we give a brief justification of why we believe that assuming that $\beta_i(t) = \pi_i(t)/(1+\hat{\tau})$ for $i = 1, 2, 3$ is a very good approximation of our full model in certain circumstances. Figure 3.3 shows simulations of the total prevalence of disease in addicts and the total prevalence of disease in needles for both the full Optimistic Model and the "addict only" approximation. We have assumed that in each model initially a fraction 0.01 of all addicts are in stage one infectivity and that a fraction 0.00935 $(= 0.01/(1+\hat{\tau}))$ of all needles are in state one infectivity. It is clear from the figure that the dynamic behaviour of these models is so similar as to be indistinguishable. While we have only illustrated

Figure 3.4: Optimistic Model when $R_0 < 1$



this for the total prevalence of disease, the same is also true of the individual components $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2$ and β_3. Other simulations (not illustrated) suggest for any parameter estimates if these models share the same initial conditions then their behaviour is virtually identical. It also appears to be true that irrespective of whether these models share the same initial conditions they tend to the same equilibrium prevalence of disease.

Parameter Set Two - $R_0 = 0.901$

We now simulate the Optimistic Model using the second set of parameter estimates where $\phi = 0.853$ which gives $R_0 = 0.901$. Figure 3.4 shows the Optimistic Model simulated over 160 years. At time zero we have assumed that the population is in an endemic steady state where 2.4% of the total population of addicts are in stage one infectivity, 30.6% are in stage two infectivity and 21.5% are in stage three infectivity. We also assume that 2.2% of the total population of needles are in state one infectivity, 28.6% are in state two infectivity and 20.1% are in state three infectivity. These values correspond to the endemic equilibrium for the Optimistic Model using the first set of parameters. We suppose that at time zero R_0 has been reduced from 2.2 down to 0.9. As in Figure 3.1 each line on the figure represents the spread of the various stages of infectivity among the addict population and needle population. It is clear from the figure that the disease dies out in all addicts and all needles and after about 150 years the Optimistic Model has almost reached the disease-free equilibrium. Other

simulations carried out for various initial conditions and parameter estimates suggest that the behaviour of our model is consistent with Theorem 3.2.

3.9 Summary of Results for the Optimistic Model

We began this chapter by discussing how to incorporate three types of infectious needles into the framework of our existing model (the Simple Model). We defined the three types of infectious needles which our model would contain and this led us on to the problem of determining addict-needle interactions, in other words determining which state a given type of needle is left in after use by a given type of addict. We argued that it is not possible to accurately assess the outcome of each addict-needle interaction and to overcome this problem we adopted a best case scenario. We picked a set of addict-needle interaction assumptions which may be more optimistic than would reasonably be expected. By optimistic we mean that under this set of assumptions we would expect the disease to travel less quickly, have a lower long term prevalence level and respond more favourably to control strategies than might occur in reality. We then derived a differential equation model which contained three types of infectious addicts and three types of infectious needles and this optimistic set of addict-needle interaction assumptions. This model represents our best case scenario.

We computed the basic reproductive number for this model and derived analytical results relating to the effect of R_0 on the long term behaviour of the model. We showed that the disease will die out in all addicts and all needles if $R_0 \leq 1$. Conversely if $R_0 > 1$ and disease is initially present then it will persist among the population indefinitely. We then showed that a simplified version of the Optimistic Model has an endemic equilibrium solution which is locally stable when $R_0 > 1$. Moreover we showed that if $\lambda\gamma$ is much larger than the other model parameters (which we expect to be the case for realistic parameter estimates) then our full model is also locally stable when $R_0 > 1$. We next discussed a number of sufficient conditions which if satisfied ensure that the prevalence of disease in the Optimistic Model tends to the unique endemic equilibrium solution. We then briefly examined a small number of simulations of our model in order to validate our previous analytical results, and moreover establish that $R_0 > 1$ (and disease initially present) is the only condition necessary for global stability of the endemic equilibrium.

Having examined thoroughly the behaviour of our best case scenario, the Optimistic

Model, we now move on to our next three stage infectivity model, the Pessimistic Model. As its name suggests this model represents a worst case scenario and we discuss this model in the following chapter.

Chapter 4

The Pessimistic Model

4.1 Introduction

We use three models to investigate the effects of incorporating a three stage infectious period into both addicts and needles. The previous chapter dealt with the first of these models, the so called Optimistic Model, this chapter deals with the second model, the Pessimistic Model. This model adopts addict-needle interaction assumptions which may be more pessimistic than would reasonably be expected. We initially discuss these pessimistic assumptions before deriving the differential equations which represent this model. Once we have derived the model we then compute an expression for the basic reproductive number and investigate the behaviour of the Pessimistic Model in a similar manner to the Simple and Optimistic Models in the previous chapters. This chapter concludes with a brief summary of the main findings.

4.2 Pessimistic Addict-Needle Interaction Assumptions

The model discussed in this chapter is referred to as the Pessimistic Model due to the pessimistic nature of the addict-needle interaction assumptions built into it. The Optimistic Model assumed that all needles are flushed with probability one (full flushing), this was justified as an optimistic assumption by using an expression for the endemic equilibrium level of disease in the Simple Model (see p.51). This expression showed that flushing decreased the value of the endemic equilibrium, hence for our Pessimistic Model we make the opposite assumption, namely that needles are never flushed.

Under the assumption of full flushing, needles can only adopt the infectious state of the last user. If we allow needles to be flushed with a probability of less than one

it is not obvious which state the needle should be left in. Full flushing means that all contents of the needle prior to use are removed during the injection process. If a needle is not fully flushed then some of the original contents of the needle will remain after use together with some of the blood from the last user. Hence intuitively the needle should have a viral load somewhere between that of the last user and the contents of the needle prior to use. However it should be noted that a value of θ (the probability that a needle is flushed in a single stage infectivity model) which is less than one does not uniquely identify a set of addict-needle interaction assumptions in a three stage infectivity framework. This is an important issue when comparing the differences between single stage infectivity models and three stage infectivity models which we discuss in Chapter 7.

We wish our model to be more pessimistic than would reasonably be expected, hence we adopt the assumption that the state of a needle after use is taken to be that of the more infectious of the state of the needle prior to use and the state of current user. As already mentioned it is generally accepted (Peterson et al., 1990, Anderson and May, 1991) that the viral load of an addict in stage one is greater than an addict in stage three which in turn is greater than an addict in stage two. Therefore we assume that a stage one addict will always leave needles in state one infectivity, a stage three addict will always leave state three, state two and uncontaminated needles in state three infectivity, and a stage two addict will always leave state two and uncontaminated needles in state two infectivity. In addition a susceptible addict cannot alter the current state of any infectious needle, this is analogous to $\theta = 0$ in single stage infectivity models.

These addict-needle interaction assumptions mean that needles can only become more infectious over use with the exclusion of cleaning. In the Optimistic Model the needle equations were unaffected by cleaning, this is not the case in the Pessimistic Model. If an addict successfully cleans a needle prior to use then the needle becomes uninfectious and therefore according to our pessimistic assumptions the needle must adopt the infectivity characteristics of the current user.

4.3 Model Derivation

The pessimistic addict-needle interaction assumptions in this model do not affect the addict equations, hence similarly to the Optimistic Model we have that

$$\frac{d\pi_1}{dt} = \Big(1-\sum_{i=1}^{3}\pi_i\Big)\lambda(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)(1-\phi)-(\mu+\delta_1)\pi_1,$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1-(\mu+\delta_2)\pi_2,$$

and $$\frac{d\pi_3}{dt} = \delta_2\pi_2-(\mu+\delta_3)\pi_3.$$

We now move on to the three equations which describe behaviour of the needles.

The number of infected state one needles at time $t+\Delta t$

$$\begin{aligned}
&= \{\text{number of state one infectious needles at time } t\}\\
&\quad +\{(\text{number of non state one needles at time } t)\\
&\quad \times(\text{fraction of needles used by stage one addicts in } [t,t+\Delta t))\}\\
&\quad -\{(\text{number of state one infected needles at time } t)\\
&\quad \times(\text{fraction of needles used and successfully cleaned prior to use}\\
&\quad \text{by non stage one addicts in } [t,t+\Delta t))\}\\
&\quad -\{\text{number of state one infectious needles exchanged in } [t,t+\Delta t)\}.
\end{aligned}$$

Thus

$$\begin{aligned}
m\beta_1(t+\Delta t) &= m\beta_1(t)+m\lambda\Delta t\gamma\pi_1(t)(1-\beta_1(t))-m\lambda\Delta t\gamma\phi(1-\pi_1(t))\beta_1(t)\\
&\quad -m\beta_1(t)\tau\Delta t+o(\Delta t).
\end{aligned}$$

Subtracting $m\beta_1(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t\to 0$ we deduce that

$$\frac{d\beta_1}{dt} = \lambda\gamma(1-\beta_1)\pi_1-\beta_1(1-\pi_1)\phi\lambda\gamma-\beta_1\tau.$$

The number of infected state two needles at time $t+\Delta t$

$$\begin{aligned}
&= \{\text{number of state two infectious needles at time } t\}\\
&\quad +\{(\text{number of uncontaminated needles at time } t)\\
&\quad \times(\text{fraction of needles used by stage two addicts in } [t,t+\Delta t))\}
\end{aligned}$$

+{(number of state three and state one needles at time t)

×(fraction of needles used and cleaned prior to use by stage two addicts in $[t, t+\Delta t)$)}

−{(number of state two infected needles at time t)

×(fraction of needles used by stage one or stage three addicts in $[t, t+\Delta t)$)}

−{(number of state two infected needles at time t)

×(fraction of needles used and cleaned prior to use by uninfected addicts in $[t, t+\Delta t)$)}

−{number of state two infectious needles exchanged in $[t, t+\Delta t)$}.

Thus

$$\begin{aligned} m\beta_2(t+\Delta t) &= m\beta_2(t) + m\lambda\Delta t\gamma\pi_2(t)\Big(1-\sum_{i=1}^{3}\beta_i(t)\Big) + m\lambda\Delta t\gamma\phi\beta_1(t)\pi_2(t) \\ &\quad + m\lambda\Delta t\gamma\phi\beta_3(t)\pi_2(t) - m\lambda\Delta t\gamma(\pi_1(t)+\pi_3(t))\beta_2(t) \\ &\quad - m\lambda\Delta t\gamma\phi\Big(1-\sum_{i=1}^{3}\pi_i(t)\Big)\beta_2(t) - m\beta_2(t)\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m\beta_2(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\begin{aligned} \frac{d\beta_2}{dt} &= \lambda\gamma\Big(1-\sum_{i=1}^{3}\beta_i\Big)\pi_2 + \beta_1\pi_2\phi\lambda\gamma + \beta_3\pi_2\phi\lambda\gamma - \beta_2\pi_3\lambda\gamma - \beta_2\pi_1\lambda\gamma \\ &\quad - \beta_2\lambda\gamma\phi\Big(1-\sum_{i=1}^{3}\pi_i\Big) - \beta_2\tau. \end{aligned}$$

The number of infected state three needles at time $t+\Delta t$

= {number of state three infectious needles at time t}

+{(number of uncontaminated and state two needles at time t)

×(fraction of needles used by stage three addicts in $[t, t+\Delta t)$)}

+{(number of state one needles at time t)

×(fraction of needles used and cleaned prior to use by stage three addicts in $[t, t+\Delta t)$)}

−{(number of state three infected needles at time t)

$\times$(fraction of needles used by stage one addicts in $[t, t+\Delta t)$)}

$-$\{(number of state three infected needles at time t)

$\times$(fraction of needles used and cleaned prior to use by

uninfected or stage two addicts in $[t, t+\Delta t)$)}

$-$\{number of state three infectious needles exchanged in $[t, t+\Delta t)$\}.

Thus

$$\begin{aligned} m\beta_3(t+\Delta t) &= m\beta_3(t) + m\lambda\Delta t\gamma\pi_3(t)(1-\beta_1(t)-\beta_3(t)) + m\lambda\Delta t\gamma\phi\beta_1(t)\pi_3(t) \\ &\quad -m\lambda\Delta t\gamma\beta_3(t)\pi_1(t) - m\beta_3(t)\lambda\Delta t\gamma\phi(1-\pi_1(t)-\pi_3(t)) - m\beta_3(t)\tau\Delta t \\ &\quad +o(\Delta t). \end{aligned}$$

Subtracting $m\beta_3(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\begin{aligned} \frac{d\beta_3}{dt} &= \lambda\gamma\pi_3(1-\beta_1-\beta_3) + \lambda\gamma\phi\beta_1\pi_3 - \lambda\gamma\beta_3\pi_1 - \beta_3\lambda\gamma\phi(1-\pi_1-\pi_3) \\ &\quad -\beta_3\tau. \end{aligned}$$

Hence the system of differential equations which describes the spread of the disease is:

$$\frac{d\pi_1}{dt} = \left(1-\sum_{i=1}^{3}\pi_i\right)\lambda(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)(1-\phi) - (\mu+\delta_1)\pi_1, \tag{4.1}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2, \tag{4.2}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3, \tag{4.3}$$

$$\frac{d\beta_1}{dt} = \lambda\gamma(1-\beta_1)\pi_1 - \beta_1(1-\pi_1)\phi\lambda\gamma - \beta_1\tau, \tag{4.4}$$

$$\begin{aligned} \frac{d\beta_2}{dt} &= \lambda\gamma\left(1-\sum_{i=1}^{3}\beta_i\right)\pi_2 + \beta_1\pi_2\phi\lambda\gamma + \beta_3\pi_2\phi\lambda\gamma - \beta_2\pi_3\lambda\gamma - \beta_2\pi_1\lambda\gamma \\ &\quad -\beta_2\lambda\gamma\phi\left(1-\sum_{i=1}^{3}\pi_i\right) - \beta_2\tau, \end{aligned} \tag{4.5}$$

and $$\begin{aligned} \frac{d\beta_3}{dt} &= \lambda\gamma\pi_3(1-\beta_1-\beta_3) + \lambda\gamma\phi\beta_1\pi_3 - \lambda\gamma\beta_3\pi_1 - \beta_3\lambda\gamma\phi(1-\pi_1-\pi_3) \\ &\quad -\beta_3\tau, \end{aligned} \tag{4.6}$$

with suitable initial conditions: $0 \le \pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$, $\beta_1(0)$, $\beta_2(0)$, $\beta_3(0)$, $\pi_1(0) + \pi_2(0) + \pi_3(0) \le 1$ and $\beta_1(0) + \beta_2(0) + \beta_3(0) \le 1$.

4.4 The Basic Reproductive Number

As with the previous models we wish to derive an explicit expression for the basic reproductive number for the model defined by eqns (4.1)-(4.6). While the Pessimistic Model has much more complex dynamic equations than the Optimistic Model the method of deriving R_0 is very similar to that for the Optimistic Model. The reason for this similarity is that the Pessimistic Model is only more structurally complicated than the Optimistic Model once an epidemic has passed the initial exponential growth stage. At the start of an epidemic both the proportions of susceptible addicts and uninfected needles in the population will be sufficiently close to one that we can ignore interactions between infectious addicts and infectious needles. In these circumstances there is little difference between the Optimistic and Pessimistic Models as we shall now show.

Again consider a single newly infectious addict entering a totally susceptible population of addicts and needles at equilibrium. We have from Section 3.6 that this single addict will on average infect

$$\frac{\lambda}{\mu+\delta_1}$$

needles during his or her entire stage one lifetime, and

$$\frac{\lambda\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}$$

needles during his or her entire stage two lifetime, and

$$\frac{\lambda\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}$$

needles during his or her entire stage three lifetime. By again dealing with each type of infectious needle separately and using the same method and notation as in Section 3.6 we have that,

$$E_1 = \frac{(1-\phi)\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected)} + \text{P(needle not flushed)}E_1\Big].$$

The Pessimistic Model assumes that a susceptible addict always leaves an infectious needle in the same infectious state as it was prior to use. In other words susceptible addicts never flush needles. Hence solving for E_1 gives

$$E_1 = \frac{(1-\phi)\alpha_1}{\hat{\tau}+\phi},$$

where $\hat{\tau} = \tau/\lambda\gamma$. Following an identical argument for state two and state three infectious needles we find that

$$E_2 = \frac{(1-\phi)\alpha_2}{\hat{\tau}+\phi},$$

and $$E_3 = \frac{(1-\phi)\alpha_3}{\hat{\tau}+\phi}.$$

We now have the expected number of addicts infected by a single state one, state two and state three infectious needle. Putting these expectations together with the expected number of each type of needle an addict creates during his or her entire infectious lifetime gives us

$$R_0 = \frac{\lambda(1-\phi)}{(\mu+\delta_1)(\hat{\tau}+\phi)}\left[\alpha_1 + \frac{\alpha_2\delta_1}{\mu+\delta_2} + \frac{\alpha_3\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right]. \tag{4.7}$$

Note that as mentioned previously this expression for R_0 is very similar to that for the Optimistic Model.

From the form of the expression in eqn (4.7) we can see that in the special case where addicts never successfully clean needles prior to use ($\phi = 0$) and needles are never exchanged for uncontaminated needles ($\tau = 0$), then $R_0 = \infty$. This is sensible since together with the assumption that addicts never flush needles we have that once a needle is contaminated it remains contaminated for all time. However this special case is unrealistic due to empirical evidence which shows that some addicts do successfully clean needles prior to use and moreover that needles must be exchanged or replaced eventually since a needle has a limited working lifetime.

4.5 Analytical Results

This section investigates the analytical properties of the Pessimistic Model. As with previous models we are mainly concerned with the global stability of the equilibrium solutions and in particular under what conditions the disease dies out or persists in the population. As in the Simple and Optimistic Models we assume that all parameters are strictly positive except for ϕ which is less than unity. As in the Optimistic Model we also assume that $\alpha_1 \geq \alpha_3 \geq \alpha_2$. The Pessimistic Model is more complex than the Simple or Optimistic Models however the proofs in this section do still follow the same general arguments as in these simpler models. In particular it is no longer trivial to show that there exists a unique endemic equilibrium when $R_0 > 1$. It is this result we deal with first.

Theorem 4.1 *If $R_0 \leq 1$ the system of equations (4.1)-(4.6) has a unique equilibrium solution where the disease has died out in both addicts and needles. If $R_0 > 1$ then there is still the equilibrium where the disease has died out, however there is also a unique endemic equilibrium.*

Proof.

Suppose that $\pi_1^\star$, $\pi_2^\star$, $\pi_3^\star$, $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ denote respectively the equilibrium values of π_1, π_2, π_3, β_1, β_2 and β_3. Let $L = 1 + (\delta_1/(\mu+\delta_2)) + (\delta_1\delta_2/(\mu+\delta_2)(\mu+\delta_3))$, define $\eta_1 = (1/L)$, $\eta_2 = (\delta_1\eta_1)/(\mu+\delta_2)$ and $\eta_3 = (\delta_2\eta_2)/(\mu+\delta_3)$. Hence from $\frac{d\pi_2}{dt} = \frac{d\pi_3}{dt} = 0$ we have that $\pi_1^\star = \eta_1\pi^\star$, $\pi_2^\star = \eta_2\pi^\star$ and $\pi_3^\star = \eta_3\pi^\star$ where $\pi^\star = \pi_1^\star + \pi_2^\star + \pi_3^\star$.

Solving $\frac{d\beta_1}{dt} = 0$ gives

$$\beta_1^\star = \frac{\pi_1^\star}{\pi_1^\star(1-\phi) + \hat{\tau} + \phi}, \tag{4.8}$$

where $\hat{\tau} = \tau/(\lambda\gamma)$.

Solving $\frac{d\beta_3}{dt} = 0$ gives

$$\beta_3^\star = \frac{\pi_3^\star(1-\beta_1^\star(1-\phi))}{\pi_1^\star(1-\phi) + \pi_3^\star(1-\phi) + \hat{\tau} + \phi}, \tag{4.9}$$

now using eqn (4.8) to replace $\beta_1^\star$ gives

$$\beta_3^\star = \frac{\pi_3^\star(\hat{\tau}+\phi)}{(\pi_1^\star(1-\phi) + \pi_3^\star(1-\phi) + \hat{\tau} + \phi)(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)}. \tag{4.10}$$

Solving $\frac{d\beta_2}{dt} = 0$ gives

$$\beta_2^\star = \frac{\pi_2^\star(1-\beta_1^\star(1-\phi) - \beta_3^\star(1-\phi))}{\pi^\star(1-\phi) + \hat{\tau} + \phi}, \tag{4.11}$$

now using eqns (4.8) and (4.10) to replace $\beta_1^\star$ and $\beta_3^\star$ respectively we find that

$$\beta_2^\star = \frac{\pi_2^\star(\hat{\tau}+\phi)}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)(\pi_1^\star(1-\phi) + \pi_3^\star(1-\phi) + \hat{\tau} + \phi)}. \tag{4.12}$$

Using η_1, η_2 and η_3 we can now write each of $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ in terms of $\pi^\star$ only. Hence we have that

$$\beta_1^\star = \frac{\eta_1\pi^\star}{\eta_1\pi^\star(1-\phi) + \hat{\tau} + \phi}, \tag{4.13}$$

$$\beta_2^\star = \frac{\eta_2\pi^\star(\hat{\tau}+\phi)}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)(\hat{\tau} + \phi + (\eta_1+\eta_3)\pi^\star(1-\phi))}, \tag{4.14}$$

and

$$\beta_3^\star = \frac{\eta_3\pi^\star(\hat{\tau}+\phi)}{(\hat{\tau} + \phi + \eta_1\pi^\star(1-\phi))(\hat{\tau} + \phi + (\eta_1+\eta_3)\pi^\star(1-\phi))}. \tag{4.15}$$

Now consider the equilibrium solution obtained by setting $\frac{d\pi_1}{dt} = 0$,

$$\beta_1^\star\alpha_1 + \beta_2^\star\alpha_2 + \beta_3^\star\alpha_3 = \frac{\eta_1\pi^\star(\mu+\delta_1)}{(1-\pi^\star)\lambda(1-\phi)}. \tag{4.16}$$

Since $\pi^\star$ is a multiplicative factor in each of $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ we have that $\pi^\star = 0$ is always a solution to eqn (4.16) and consequently there always exists a disease-free equilibrium solution. The other (non zero) solutions must satisfy

$$\begin{aligned}\frac{\eta_1(\mu+\delta_1)}{(1-\pi^\star)\lambda(1-\phi)} &= \frac{\alpha_1\eta_1}{\eta_1\pi^\star(1-\phi)+\hat{\tau}+\phi} \\ &+\frac{\eta_2(\hat{\tau}+\phi)\alpha_2}{(\hat{\tau}+\phi+\pi^\star(1-\phi))(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))} \\ &+\frac{\eta_3(\hat{\tau}+\phi)\alpha_3}{(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))(\hat{\tau}+\phi+\eta_1\pi^\star(1-\phi))}. \quad (4.17)\end{aligned}$$

We wish to show that eqn (4.17) has only a single root in $(0,1)$ when $R_0 > 1$ and no roots in $(0,1)$ when $R_0 \leq 1$ and hence we have a unique endemic solution when $R_0 > 1$ and no endemic solution when $R_0 \leq 1$. We do not attempt to derive the roots explicitly but instead focus on the conditions necessary for the existence of only a single root in $(0,1)$. Let

$$F(\pi^\star) = \frac{\eta_1(\mu+\delta_1)}{(1-\pi^\star)\lambda(1-\phi)}. \quad (4.18)$$

It is clear that $F(\pi^\star)$ is strictly monotone increasing for $\pi^\star \in (0,1)$. We have that

$$F(0) = \frac{\eta_1(\mu+\delta_1)}{\lambda(1-\phi)},$$

and $\lim_{\pi^\star\to 1} F(\pi^\star) = \infty$. Let

$$\begin{aligned}G(\pi^\star) &= \frac{\alpha_1\eta_1}{\eta_1\pi^\star(1-\phi)+\hat{\tau}+\phi} \\ &+ \frac{(\hat{\tau}+\phi)\alpha_2\eta_2}{(\hat{\tau}+\phi+\pi^\star(1-\phi))(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))} \\ &+ \frac{(\hat{\tau}+\phi)\alpha_3\eta_3}{(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))(\hat{\tau}+\phi+\eta_1\pi^\star(1-\phi))}.\end{aligned}$$

Each term in $G(\pi^\star)$ is strictly monotone decreasing in $\pi^\star$ for $\pi^\star \geq 0$. Hence $G(\pi^\star)$ is strictly monotone decreasing and $\lim_{\pi^\star\to\infty} G(\pi^\star) = 0$. So $F(\pi^\star)$ is strictly monotone increasing in $(0,1)$ and $G(\pi^\star)$ is strictly monotone decreasing in $(0,1)$. Moreover for ϵ small and positive we have that $F(1-\epsilon) > G(1-\epsilon)$.

Now consider the initial conditions of F and G, we have three distinct cases, firstly if $G(0) > F(0)$ then then the two functions must cross in $(0,1)$ and hence eqn (4.17) has a unique strictly positive root, $\pi^\star \in (0,1)$. Secondly if $G(0) < F(0)$ then the functions never cross in $(0,1)$ and eqn (4.17) has no root in $(0,1)$. Finally if $G(0) = F(0)$ eqn

(4.17) has a single root $\pi^\star = 0$. Hence we are particularly interested in the first case where $G(0) > F(0)$ as this condition gives rise to a unique strictly positive solution. We now show that for this case not only is $\pi^\star \in (0,1)$ but also $\beta^\star \in (0,1)$ and hence we have a unique feasible endemic solution.

Using eqns (4.8), (4.10) and (4.12) we have that

$$\beta_1^\star + \beta_3^\star = \frac{\pi_1^\star + \pi_3^\star}{(\pi_1^\star + \pi_3^\star)(1-\phi) + \hat{\tau} + \phi}, \tag{4.19}$$

and

$$\beta^\star = \frac{\pi^\star}{\pi^\star(1-\phi) + \hat{\tau} + \phi}. \tag{4.20}$$

Therefore it follows immediately that $\beta^\star \in (0,1)$ if $\pi^\star \in (0,1)$, thus when $G(0) > F(0)$ we have that $\pi^\star \in (0,1)$ and $\beta^\star \in (0,1)$ and hence we have a unique endemic solution when $(G(0)/F(0)) > 1$, where

$$G(0) = \frac{\alpha_1\eta_1}{\hat{\tau}+\phi} + \frac{\alpha_2\eta_2(\hat{\tau}+\phi)}{(\hat{\tau}+\phi)^2} + \frac{\alpha_3\eta_3(\hat{\tau}+\phi)}{(\hat{\tau}+\phi)^2},$$

and

$$F(0) = \frac{\eta_1(\mu+\delta_1)}{\lambda(1-\phi)}.$$

Hence

$$\frac{G(0)}{F(0)} = \frac{\lambda(1-\phi)}{\eta_1(\mu+\delta_1)(\hat{\tau}+\phi)}(\alpha_1\eta_1 + \alpha_2\eta_2 + \alpha_3\eta_3),$$

which simplifies to R_0. Hence we have shown that if $R_0 > 1$ there is a unique endemic equilibrium $(\pi^\star, \beta^\star)$ with $\pi^\star > 0$ and $\beta^\star > 0$ in addition to the disease-free one, whilst if $R_0 \leq 1$ there is only the disease-free equilibrium. This completes the proof of Theorem 4.1.•

Theorem 4.2 *If $R_0 \leq 1$ then whatever the initial state the disease will die out in both addicts and needles.*

Proof.

This proof is similar in structure to that of Theorem 2.2. Eqns (4.1)-(4.3) are the same as eqns (2.1)-(2.3), hence as in the Optimistic Model we can use Lemma 2.1 and Corollary 2.1 directly. The needle equations in the Pessimistic Model contain more non-linearities than the needle equations in the Optimistic Model, however we can bound these more complex equations above by simple linear forms. From eqns (4.4)-(4.6) we have that

$$\frac{d\beta_1}{dt} = (1-\beta_1(1-\phi))\lambda\gamma\pi_1 - \beta_1(\lambda\gamma\phi + \tau),$$

$$\leq \lambda\gamma\pi_1 - \beta_1(\lambda\gamma\phi + \tau), \tag{4.21}$$

$$\frac{d\beta_2}{dt} = (1-\beta(1-\phi))\lambda\gamma\pi_2 - \beta_2(1-\phi)\lambda\gamma\pi_3 - \beta_2(1-\phi)\lambda\gamma\pi_1 - \beta_2(\lambda\gamma\phi+\tau),$$

$$\leq \lambda\gamma\pi_2 - \beta_2(\lambda\gamma\phi + \tau), \tag{4.22}$$

and $$\frac{d\beta_3}{dt} = (1-\beta_1(1-\phi) - \beta_3(1-\phi))\lambda\gamma\pi_3 - \beta_3(1-\phi)\lambda\gamma\pi_1 - \beta_3(\lambda\gamma\phi+\tau),$$

$$\leq \lambda\gamma\pi_3 - \beta_3(\lambda\gamma\phi + \tau). \tag{4.23}$$

Hence we can express each needle equation in a form similar to that of the addict equations, therefore it follows similarly to the proof of Lemma 2.1 that

$$\beta_1^\infty \leq \frac{\pi_1^\infty}{\hat{\tau}+\phi},$$

$$\beta_2^\infty \leq \frac{\pi_2^\infty}{\hat{\tau}+\phi} \leq \frac{\delta_1\pi_1^\infty}{(\mu+\delta_2)(\hat{\tau}+\phi)}, \qquad \text{using Lemma 2.1,}$$

and $$\beta_3^\infty \leq \frac{\pi_3^\infty}{\hat{\tau}+\phi} \leq \frac{\delta_1\delta_2\pi_1^\infty}{(\mu+\delta_2)(\mu+\delta_3)(\hat{\tau}+\phi)}, \qquad \text{using Corollary 2.1.}$$

Suppose first that $\pi_1^\infty > 0$. Then from eqn (4.1) we have, given $\epsilon > 0$,

$$\frac{d\pi_1}{dt} \leq (1-\pi_1)\lambda(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1-\phi) - (\mu+\delta_1)\pi_1,$$

$$\leq (1-\pi_1)\lambda(1-\phi)(\beta_1^\infty\alpha_1 + \beta_2^\infty\alpha_2 + \beta_3^\infty\alpha_3 + \epsilon) - (\mu+\delta_1)\pi_1, \qquad \forall t \geq t_1(\epsilon),$$

$$\leq (1-\pi_1)\frac{\lambda(1-\phi)}{\phi+\hat{\tau}}\left[\alpha_1 + \frac{\delta_1\alpha_2}{\mu+\delta_2} + \frac{\delta_1\delta_2\alpha_3}{(\mu+\delta_2)(\mu+\delta_3)} + \epsilon_1\right]\pi_1^\infty - (\mu+\delta_1)\pi_1,$$

$$\text{where } \epsilon_1 = \frac{\epsilon(\hat{\tau}+\phi)}{\pi_1^\infty},$$

$$\leq (1-\pi_1)(\mu+\delta_1)(R_0+\epsilon_2)\pi_1^\infty - (\mu+\delta_1)\pi_1, \qquad \text{where } \epsilon_2 = \frac{\epsilon_1\lambda(1-\phi)}{(\mu+\delta_1)(\hat{\tau}+\phi)},$$

$$\leq (\mu+\delta_1)[(R_0+\epsilon_2)\pi_1^\infty - \pi_1(1+R_0\pi_1^\infty)].$$

As for the Optimistic Model the result follows directly using the latter part of the proof of Theorem 2.2.•

Theorem 4.3 *If $R_0 > 1$ then there is still the equilibrium where the disease has died out and this equilibrium is unstable.*

Proof.

Consider the linearised system of eqns (4.1)-(4.6), evaluated at the disease-free equilibrium. This system can be represented in matrix form as

$$\frac{d\mathbf{x}}{dt} = \mathbf{J}\mathbf{x},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$ and

$$\mathbf{J} = \begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda\alpha_1(1-\phi) & \lambda\alpha_2(1-\phi) & \lambda\alpha_3(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\ \lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\tau) & 0 & 0 \\ 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\tau) & 0 \\ 0 & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\tau) \end{bmatrix}.$$

We wish to show that at least one eigenvalue of J has a strictly positive real part. Using the Routh-Hurwitz conditions it is sufficient to show that the constant term, a_6, in the characteristic equation of J,

$$\omega^6 + a_1\omega^5 + a_2\omega^4 + a_3\omega^3 + a_4\omega^2 + a_5\omega + a_6 = 0$$

is strictly negative. By substituting $\lambda\gamma\phi + \tau$ for $\lambda\gamma + \tau$ in the proof of Theorem 3.3 it is straightforward to show that

$$\begin{aligned} a_6 &= (\lambda\gamma\phi+\tau)^3(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3) \\ &\quad \times \left[1 - \frac{\lambda^2\gamma\alpha_1(1-\phi)(\mu+\delta_2)(\mu+\delta_3) + \lambda^2\gamma\alpha_2(1-\phi)\delta_1(\mu+\delta_3)}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma\phi+\tau)}\right. \\ &\qquad \left. - \frac{\lambda^2\gamma\alpha_3(1-\phi)\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(\lambda\gamma\phi+\tau)}\right]. \end{aligned}$$

Substituting in the expression for R_0 from eqn (4.7) we find that

$$a_6 = (\lambda\gamma\phi+\tau)^3(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(1-R_0),$$

hence if $R_0 > 1$ then $a_6 < 0$ and the result follows.•

Theorem 4.4 *If $R_0 > 1$ and either $\pi(0) > 0$ or $\beta(0) > 0$ then there exists a fixed $\epsilon > 0$ depending only on the model parameters and not the initial conditions such that for some $T^+ > 0$*

$$\pi_1 \geq \epsilon\pi_1^\star,\ \pi_2 \geq \epsilon\pi_2^\star,\ \pi_3 \geq \epsilon\pi_3^\star \text{ and } \beta_1 \geq \epsilon\beta_1^\star, \quad \forall t \geq T^+. \tag{4.24}$$

Proof.

The proof follows a similar method to the equivalent result the Simple and Optimistic Models. As previously we can use Lemma 2.2 and Corollary 2.3 directly, hence

$$\pi_{2,\infty} \geq \frac{\delta_1 \pi_{1,\infty}}{\mu + \delta_2},$$

and

$$\pi_{3,\infty} \geq \frac{\delta_1 \delta_2 \pi_{1,\infty}}{(\mu + \delta_2)(\mu + \delta_3)}.$$

Using eqn (4.4) we have that

$$\begin{aligned} \frac{d\beta_1}{dt} &= \lambda\gamma\pi_1 - \beta_1(\lambda\gamma\phi + \tau) - \pi_1\lambda\gamma\beta_1(1-\phi), \\ &\geq \lambda\gamma\pi_1 - \beta_1(\lambda\gamma\phi + \tau) - \lambda\gamma\beta_1(1-\phi), \\ &= \lambda\gamma\pi_1 - \beta_1(\lambda\gamma + \tau). \end{aligned}$$

Therefore

$$\frac{d}{dt}\left[\beta_1 \exp[(\lambda\gamma + \tau)t]\right] \geq \pi_1 \lambda\gamma \exp[(\lambda\gamma + \tau)t],$$

and from the form of this, arguing as in the proof of Lemma 2.2 we have that

$$\beta_{1,\infty} \geq \frac{\lambda\gamma\pi_{1,\infty}}{\lambda\gamma + \tau}.$$

As previously from the above results is it sufficient to show that $\pi_{1,\infty} > 0$ since for t sufficiently large $\pi_1(t) \geq \pi_{1,\infty} - \nu$ where ν is arbitrarily small.

Lemma 4.1 *Provided at least one of* $\pi_1(t)$, $\pi_2(t)$, $\pi_3(t)$, $\beta_1(t)$, $\beta_2(t)$ *and* $\beta_3(t)$ *is strictly positive at* $t = 0$ *then* $\pi_1(\Delta t) > 0$, $\pi_2(\Delta t) > 0$, $\pi_3(\Delta t) > 0$, $\beta_1(\Delta t) > 0$, $\beta_2(\Delta t) > 0$, *and* $\beta_3(\Delta t) > 0$ *for* Δt *small and positive.*

Proof.

Let $\psi = 1 - \pi$, then we have that

$$\begin{aligned} \frac{d\pi}{dt} &= (1-\pi)\lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3) - \mu\pi - \delta_3\pi_3, \\ \frac{d\psi}{dt} &= -\lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)\psi + \mu(1-\psi) + \delta_3\pi_3, \end{aligned}$$

and

$$\frac{d\beta}{dt} = \lambda\gamma\pi - \beta(\lambda\gamma\pi(1-\phi) + \phi\lambda\gamma + \tau).$$

We need to consider four separate initial conditions:

1. Suppose that initially $\beta(0) = 0$ which implies that $\pi(0) > 0$. We have that

$$\beta(\Delta t) = \pi(0)\lambda\gamma\Delta t + o(\Delta t) > 0$$

and $\pi(\Delta t) = \pi(0) - (\mu\pi(0) + \delta_3\pi_3(0))\Delta t + o(\Delta t) > 0$ (for Δt small).

If $\pi(0) < 1$ this implies that $\psi(0) > 0$, hence $\psi(\Delta t) > 0$ for small enough Δt. If $\pi(0) = 1$ then $\psi(0) = 1$ and

$$\psi(\Delta t) \geq \mu\pi(0)\Delta t + o(\Delta t) > 0.$$

Hence by choosing Δt small enough and starting at $t = \Delta t$ we can if necessary assume that $\pi(0) > 0$, $\psi(0) > 0$ and $\beta(0) > 0$. If $\pi_1(0) = 0$ then $\pi_1(\Delta t) = \psi(0)\lambda(1-\phi)\hat{\beta}(0)\Delta t + o(\Delta t) > 0$, hence we can additionally assume that $\pi_1(0) > 0$. If $\pi_2(0) = 0$ then $\pi_2(\Delta t) = \delta_1\pi_1(0)\Delta t + o(\Delta t) > 0$, hence we can also assume that $\pi_2(0) > 0$. Similarly if $\pi_3(0) = 0$ then $\pi_3(\Delta t) = \delta_2\pi_2(0)\Delta t + o(\Delta t) > 0$ and we can moreover assume that $\pi_3(0) > 0$. Hence we can assume that $\beta(0)$, $\pi_1(0)$, $\pi_2(0)$ and $\pi_3(0)$ are all strictly positive. Now if $\beta_1(0) = 0$ then $\beta_1(\Delta t) = \lambda\gamma\pi_1(0)\Delta t + o(\Delta t) > 0$ so we can additionally assume that $\beta_1(0) > 0$. If $\beta_3(0) = 0$ then

$$\beta_3(\Delta t) = (1-\beta_1(0)(1-\phi))\lambda\gamma\pi_3(0) + o(\Delta t) > 0 \quad \Big(\text{provided that } \beta_1(0)(1-\phi) \neq 1\Big).$$

Suppose that $\beta_1(0) = 1$ then

$$\beta_1(\Delta t) = 1 - (\phi\lambda\gamma(1-\pi_1(0)) + \tau)\Delta t + o(\Delta t) < 1,$$

hence we can always ensure that $\beta_1(0) < 1$ by moving the origin to $t = \Delta t$. Thus for Δt small we can assume that $\beta_1(0)(1-\phi) < 1$ (since $\beta_1(1-\phi) < \beta_1$). So now we have that $\pi_i(0) > 0$ for $i = 1, 2, 3$ and $\beta_j > 0$ for $j = 1, 3$. If $\beta_2(0) = 0$ then

$$\beta_2(\Delta t) = (1 - \beta_{1+3}(0)(1-\phi))\lambda\gamma\pi_2(0)\Delta t + o(\Delta t) > 0$$

$$\Big(\text{provided that } \beta_{1+3}(0)(1-\phi) \neq 1\Big).$$

Suppose that $\beta_{1+3}(0) = 1$. Now since

$$\frac{d\beta_{1+3}}{dt} = \lambda\gamma\pi_{1+3}(1-\beta_{1+3}) - \beta_{1+3}(1-\pi_{1+3})\lambda\gamma\phi - \beta_{1+3}\tau,$$

then

$$\beta_{1+3}(\Delta t) = 1 - \Big(\lambda\gamma\phi(1-\pi_{1+3}(0)) + \tau\Big)\Delta t + o(\Delta t) < 1, \qquad \text{for small } \Delta t.$$

Therefore by shifting the origin to $t = \Delta t$ for Δt small we can always ensure that $\beta_{1+3}(0) < 1$ which implies that $\beta_{1+3}(0)(1-\phi) \neq 1$. Hence if initially $\beta(0) = 0$ and $\pi(0) > 0$ then all components are strictly positive at $t = \Delta t$ for Δt small.

2. Suppose that $\pi(0) = 0$, which implies that $\psi(0) = 1$ and $\beta(0) > 0$. We have that at $t = 0$,

$$\frac{d\pi}{dt} = \lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3),$$

and $$\frac{d\beta}{dt} = -\beta(\lambda\gamma\phi + \tau).$$

Hence

$$\pi(\Delta t) = \lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)\Delta t + o(\Delta t) > 0,$$

$$\beta(\Delta t) = \beta(0) - \beta(0)(\lambda\gamma\phi + \tau)\Delta t + o(\Delta t) > 0, \qquad \text{for } \Delta t \text{ small,}$$

and $$\psi(\Delta t) = \psi(0) - \lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)\Delta t + o(\Delta t) > 0,$$

for Δt small.

Hence we can assume that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$ and by Case 1 we have that all components are strictly positive at $t = \Delta t$ for Δt small and positive.

3. Suppose that $\beta(0) > 0$, $\pi(0) > 0$, $\psi(0) > 0$, then by Case 1 we have directly that all components are strictly positive at $t = \Delta t$ for Δt small and positive.

4. Suppose that $\beta(0) > 0$, $\psi(0) = 0$ and $\pi(0) = 1$. We have that $\beta(\Delta t) > 0$ and $\pi(\Delta t) > 0$ by continuity. When $\pi(0) = 1$ we have that $d\pi/dt$ evaluated at $t = 0$ is negative and hence for $t = \Delta t$ where Δt is small and positive we can assume that $\psi(\Delta t) > 0$ therefore by shifting the origin to $t = \Delta t$ we are in Case 3 and hence we can assume all components are strictly positive.

This completes the proof of Lemma 4.1.•

From Lemma 4.1 we have that there exists fixed ϵ where $1 > \epsilon > 0$ such that if Δt is small enough $\pi_i(\Delta t) \geq \epsilon\pi_i^\star$ and $\beta_i(\Delta t) \geq \epsilon\beta_i^\star$ for $i = 1, 2, 3$. As in the Optimistic Model we must have that either $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$ or $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$. In the case where $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$ then (4.24) follows directly. Next suppose that $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$ in which case there exists $\zeta \geq \Delta t$ where $\pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star$. Let $t_0 = \inf\{\zeta \geq \Delta t, \pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star\}$ and $t_1 = \inf\{\zeta \geq t_0, \pi_1(\zeta) > \frac{1}{2}\epsilon\pi_1^\star\}$ where ϵ is fixed and positive. By the definition of t_0 we have that $\pi_1(t_0 + \nu) < \frac{1}{2}\epsilon\pi_1^\star$ if ν is small and positive. As in previous models by continuity $\pi_1(t_0) = \pi_1(t_1) = \frac{1}{2}\epsilon\pi_1^\star$, and therefore π_1 is less than $\frac{1}{2}\epsilon\pi_1^\star$ in (t_0, t_1) and greater than $\frac{1}{2}\epsilon\pi_1^\star$ just after t_1. We now show that if π_1 becomes small and remains small then all components eventually become small.

We have that in $[t_0, t_1]$, $\pi_1 \leq \frac{1}{2}\epsilon\pi_1^\star$, by again exploiting the similarities between the Optimistic and Pessimistic Models we can use Lemma 2.4 and Corollary 2.5 directly, hence we have that

$$\pi_2 \leq \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \qquad \text{for } t \geq t_0 + \bar{T}_1,$$

$$\text{and} \qquad \pi_3 \leq \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_2,$$

where $\bar{T}_1$ and $\bar{T}_2$ depend only on ϵ, Δ and the model parameters. From eqn (4.4)

$$\frac{d\beta_1}{dt} \leq \frac{1}{2}\lambda\gamma\epsilon\pi_1^\star - \beta_1(\lambda\gamma\phi + \tau), \qquad \text{in } [t_0, t_1],$$

and arguing as in the proof of Lemma 2.4,

$$\beta_1(t) \leq \frac{\left(\frac{1}{2}+\Delta\right)\pi_1^\star\epsilon}{\hat{\tau}+\phi} \qquad \text{(for } t \text{ sufficiently large, say } t \geq t_0 + \bar{T}_3\text{)}.$$

However we cannot replace $\pi_1^\star/(\hat{\tau}+\phi)$ by $\beta_1^\star$ since

$$\beta_1^\star = \frac{\pi_1^\star}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi} \leq \frac{\pi_1^\star}{\hat{\tau}+\phi}.$$

Letting

$$\epsilon_1 = \frac{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)\epsilon}{\hat{\tau}+\phi} > \epsilon,$$

we have that $\beta_1(t) \leq \left(\frac{1}{2}+\Delta\right)\epsilon_1\beta_1^\star$ if t is sufficiently large.

From eqn (4.5) we have that for $t \geq t_0 + \bar{T}_1$,

$$\frac{d\beta_2}{dt} \leq \left(\frac{1}{2}+\Delta\right)\lambda\gamma\epsilon\pi_2^\star - (\lambda\gamma\phi+\tau)\beta_2.$$

Arguing again as in Lemma 2.4 but integrating instead over $[t_0+\bar{T}_1, t]$ we find that

$$\beta_2(t) \leq \frac{\left(\frac{1}{2}+2\Delta\right)\pi_2^\star\epsilon}{\hat{\tau}+\phi}, \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_4,$$

$$= \left(\frac{1}{2}+2\Delta\right)\epsilon_2\beta_2^\star, \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_4,$$

$$\text{where} \qquad \epsilon_2 = \frac{(\pi^*(1-\phi)+\hat{\tau}+\phi)(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)\epsilon}{(\hat{\tau}+\phi)^2} > \epsilon \text{ using eqn (4.14).}$$

From eqn (4.6) we have that

$$\frac{d\beta_3}{dt} \leq \left(\frac{1}{2}+2\Delta\right)\lambda\gamma\epsilon\pi_3^\star - (\lambda\gamma\phi+\tau)\beta_3, \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_2.$$

Arguing as above but integrating over $[t_0+\bar{T}_1+\bar{T}_2, t]$ we find that $\beta_3(t) \leq \left(\frac{1}{2}+3\Delta\right)\epsilon_3\beta_3^\star$ for $t \geq t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_5$ where

$$\epsilon_3 = \frac{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)\epsilon}{(\hat{\tau}+\phi)^2} > \epsilon, \quad \text{using eqn (4.15).}$$

Hence we have that

$$\begin{aligned}\beta_{1+3} &\leq \left(\frac{1}{2}+\Delta\right)\epsilon_1\beta_1^\star + \left(\frac{1}{2}+3\Delta\right)\epsilon_3\beta_3^\star, \\ &\leq \left(\frac{1}{2}+3\Delta\right)\epsilon_4\beta_{1+3}^\star, \qquad \text{for all } t \geq t_0 + \max(\bar{T}_3, \bar{T}_1+\bar{T}_2+\bar{T}_5),\end{aligned}$$

where $\epsilon_4 = \max(\epsilon_1, \epsilon_3)$. Similarly we have that $\beta \leq \left(\frac{1}{2}+3\Delta\right)\epsilon_5\beta^\star$ for all $t \geq t_0 + \max(\bar{T}_3, \bar{T}_1+\bar{T}_2+\bar{T}_5, \bar{T}_1+\bar{T}_4)$, where $\epsilon_5 = \max(\epsilon_2, \epsilon_4)$.

Now either π_1 is below $\frac{1}{2}\epsilon\pi_1^\star$ long enough for all components to become small or π_1 increases past $\frac{1}{2}\epsilon\pi_1^\star$ before all components become small. Hence we have either that

$$(i) \quad t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1+\bar{T}_2, \bar{T}_3, \bar{T}_1+\bar{T}_4, \bar{T}_1+\bar{T}_2+\bar{T}_5],$$

or $$(ii) \quad t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1+\bar{T}_2, \bar{T}_3, \bar{T}_1+\bar{T}_4, \bar{T}_1+\bar{T}_2+\bar{T}_5].$$

We wish to show that $t_1 < T$ where T is a fixed finite value dependent only on the model parameters, ϵ and Δ. In case (ii) we have already have this result, we now show that this case must always be true by obtaining an upper bound for t_1 in case (i). As in the previous models we use the fact that the disease-free equilibrium is unstable when $R_0 > 1$ to show that π_1 cannot become arbitrarily small.

Lemma 4.2 *If $\pi_1(t)$ drops to below $\frac{1}{2}\epsilon\pi_1^\star$ at time t_0 then $\pi_1(t)$ returns to $\frac{1}{2}\epsilon\pi_1^\star$ by at least time $t_1^+ = t_0 + \max[\bar{T}_1, \bar{T}_1+\bar{T}_2, \bar{T}_3, \bar{T}_1+\bar{T}_4, \bar{T}_1+\bar{T}_2+\bar{T}_5, t_2+\bar{T}_6]$ where $t_1^+ - t_0$ is finite and depends only on Δ, ϵ and the model parameters.*

Proof.

Consider the coordinate system $\mathbf{x}'^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta, \beta_{1+3})$:

$$\begin{aligned}\frac{d\pi_1}{dt} &= (1-\pi)\lambda(\beta_1(\alpha_1-\alpha_3)+\beta\alpha_2+\beta_{1+3}(\alpha_3-\alpha_2))(1-\phi) - (\mu+\delta_1)\pi_1, \\ \frac{d\pi_2}{dt} &= \pi_1\delta_1 - (\mu+\delta_2)\pi_2, \\ \frac{d\pi_3}{dt} &= \pi_2\delta_2 - (\mu+\delta_3)\pi_3, \\ \frac{d\beta_1}{dt} &= \lambda\gamma\pi_1 - (\lambda\gamma\pi_1(1-\phi)+\lambda\gamma\phi+\tau)\beta_1, \\ \frac{d\beta}{dt} &= \lambda\gamma\pi - (\lambda\gamma\pi(1-\phi)+\lambda\gamma\phi+\tau)\beta, \\ \text{and} \quad \frac{d\beta_{1+3}}{dt} &= \lambda\gamma\pi_{1+3} - (\lambda\gamma\pi_{1+3}(1-\phi)+\lambda\gamma\phi+\tau)\beta_{1+3}.\end{aligned}$$

The Jacobian of this system at the disease-free equilibrium is $\mathbf{J} =$

$$\begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda(\alpha_1-\alpha_3)(1-\phi) & \lambda\alpha_2(1-\phi) & \lambda(\alpha_3-\alpha_2)(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\ \lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\tau) & 0 & 0 \\ \lambda\gamma & \lambda\gamma & \lambda\gamma & 0 & -(\lambda\gamma\phi+\tau) & 0 \\ \lambda\gamma & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\tau) \end{bmatrix}.$$

If M is large enough then $\mathbf{J} + M\mathbf{I}$ is a non-negative irreducible matrix and hence the characteristic equation has a simple root equal to its spectral radius (Lemma 2.1 in Nold (1980)). If the eigenvalues of $\mathbf{J}$ are ω_1, ω_2, ω_3, ω_4, ω_5 and ω_6 then the eigenvalues of $\mathbf{J}+M\mathbf{I}$ are $M+\omega_1$, $M+\omega_2$, $M+\omega_3$, $M+\omega_4$, $M+\omega_5$ and $M+\omega_6$. Hence if $M+\omega_1$ is the spectral radius of $\mathbf{J} + M\mathbf{I}$ then ω_1 is real and has a larger real part than any other eigenvalue. Now consider the previous linearised system perturbed slightly from the disease-free equilibrium. Arguing as in the proof of Lemma 3.2, the corresponding result for the Optimistic Model, given $\epsilon_2 > 0$ we can choose ϵ small enough so that

$$\frac{1}{2}\epsilon\pi_1^\star + \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star + \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star < \epsilon_2.$$

Hence for $t_1 > t \geq t_0 + \bar{T}_1 + \bar{T}_2$ we have that $\pi_1 + \pi_2 + \pi_3 < \epsilon_2$, using Lemma 2.4 and Corollary 2.5. Hence after a sufficient duration we have that $\pi < \epsilon_2$ and hence $\pi_1, \pi_{1+3} < \epsilon_2$. Therefore we have that

$$\begin{aligned} \frac{d\pi_1}{dt} &\geq (1-\epsilon_2)\lambda(\beta_1(\alpha_1-\alpha_3)+\beta_2\alpha_2+\beta_{1+3}(\alpha_3-\alpha_2))(1-\phi)-(\mu+\delta_1)\pi_1, \\ \frac{d\pi_2}{dt} &= \pi_1\delta_1-(\mu+\delta_2)\pi_2, \\ \frac{d\pi_3}{dt} &= \pi_2\delta_2-(\mu+\delta_3)\pi_3, \\ \frac{d\beta_1}{dt} &\geq \lambda\gamma\pi_1-(\lambda\gamma\epsilon_2(1-\phi)+\lambda\gamma\phi+\tau)\beta_1, \\ \frac{d\beta}{dt} &\geq \lambda\gamma\pi-(\lambda\gamma\epsilon_2(1-\phi)+\lambda\gamma\phi+\tau)\beta, \\ \text{and} \quad \frac{d\beta_{1+3}}{dt} &\geq \lambda\gamma\pi_{1+3}-(\lambda\gamma\epsilon_2(1-\phi)+\lambda\gamma\phi+\tau)\beta_{1+3}. \end{aligned}$$

In matrix form we now have $\frac{d\mathbf{x}'}{dt} \geq \mathbf{J}(\epsilon_2)\mathbf{x}'$ where $\mathbf{J}(\epsilon_2) =$

$$\begin{bmatrix}
-(\mu+\delta_1) & 0 & 0 & \lambda(\alpha_1-\alpha_3)(1-\phi)(1-\epsilon_2) & \lambda\alpha_2(1-\phi)(1-\epsilon_2) & \lambda(\alpha_3-\alpha_2)(1-\phi)(1-\epsilon_2) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\
\lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\lambda\gamma(1-\phi)\epsilon_2+\tau) & 0 & 0 \\
\lambda\gamma & \lambda\gamma & \lambda\gamma & 0 & -(\lambda\gamma\phi+\lambda\gamma(1-\phi)\epsilon_2+\tau) & 0 \\
\lambda\gamma & 0 & \lambda\gamma & 0 & 0 & -(\lambda\gamma\phi+\lambda\gamma(1-\phi)\epsilon_2+\tau)
\end{bmatrix}.$$

Denote the eigenvalues of $\mathbf{J}(\epsilon_2)$ by $\omega_1(\epsilon_2)$, $\omega_2(\epsilon_2)$, $\omega_3(\epsilon_2)$, $\omega_4(\epsilon_2)$, $\omega_5(\epsilon_2)$ and $\omega_6(\epsilon_2)$. Hence using a similar argument to previously, we have that $\omega_1(\epsilon_2)$ is real and all other eigenvalues have strictly smaller real parts. Moreover from Corollary 2.7 we have that the roots of the characteristic equation of $\mathbf{J}(\epsilon_2)$ are continuous functions of ϵ_2, hence $\omega_1(\epsilon_2) \to \omega_1(0)$ as $\epsilon_2 \to 0$. We know that from Theorem 4.3 the disease-free equilibrium of the Pessimistic Model is unstable when $R_0 > 1$, therefore $\omega_1(0)$ is strictly positive. Therefore by choosing ϵ_2 small enough we can ensure that $\omega_1(\epsilon_2) > 0$. Without loss of generality we can assume that $1 > \epsilon_2 > 0$.

As we have previously argued, for $t_1 > t \geq t_0+\bar{T}_1+\bar{T}_2$ we have that $\pi_1+\pi_2+\pi_3 < \epsilon_2$. Let $t_2 = \inf\{\zeta : \text{ for } t_1 > t \geq t_0+\zeta, \pi(t) < \epsilon_2\}$, hence if $t_2 > 0$ then $\pi(t_0+t_2) = \epsilon_2$ and t_0+t_2 is the last time before t_1 that $\pi(t) \geq \epsilon_2$, and note that $t_2 \leq \bar{T}_1+\bar{T}_2$. If $t_1 < t_0+\bar{T}_1+\bar{T}_2$ then we have the desired result. Now we consider the case where $t_1 \geq t \geq t_0+\bar{T}_1+\bar{T}_2$. We have that

$$\frac{d\mathbf{x}}{dt} \geq \mathbf{J}(\epsilon_2)\mathbf{x},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta, \beta_{1+3})$. From Lemma 2.1 in Nold (1980) $\mathbf{J}(\epsilon_2)$ has a strictly positive left eigenvector, $\mathbf{e} = (e_1, e_2, e_3, e_4, e_5, e_6)$ corresponding to its spectral radius $w_1(\epsilon_2)$. Hence

$$\mathbf{e}\frac{d\mathbf{x}}{dt} \geq \mathbf{e}\,\mathbf{J}(\epsilon_2)\mathbf{x} = w_1(\epsilon_2)\,\mathbf{e}.\mathbf{x}.$$

Thus

$$\begin{aligned}
\mathbf{e}.\mathbf{x}(t) &\geq \mathbf{e}.\mathbf{x}(t_0+t_2)\exp[w_1(\epsilon_2)(t-t_0-t_2)], \quad \text{(integrating over}[t_0+t_2, t]), \\
&\geq (e_1\pi_1+e_2\pi_2+e_3\pi_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)], \\
&\geq \pi(t_0+t_2)\min(e_1,e_2,e_3)\exp[w_1(\epsilon_2)(t-t_0-t_2)],
\end{aligned}$$

$$\begin{cases} = & \epsilon_2 \min(e_1, e_2, e_3) \exp[w_1(\epsilon_2)(t - t_0 - t_2)], & \text{if } t_2 > 0, \\ \geq & \frac{1}{2}\epsilon\pi_1^\star \min(e_1, e_2, e_3) \exp[w_1(\epsilon_2)(t - t_0 - t_2)], & \text{if } t_2 = 0. \end{cases}$$

Therefore after a time $t_0 + t_2 + \bar{T}_6$ we have that

$$\mathbf{e.x}(t) \quad > \quad \mathbf{e}\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+\Delta\right)\epsilon_1\beta_1^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_5\beta^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_4\beta_{1+3}^\star\right),$$

where $\bar{T}_6$ depends only on ϵ, Δ and the model parameters. We already know that provided $t_0 \leq t \leq t_1$ then after a time $t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5]$,

$$\mathbf{e.x}(t) \quad \leq \quad \mathbf{e}\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+\Delta\right)\epsilon_1\beta_1^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_5\beta^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_4\beta_{1+3}^\star\right).$$

However if $t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, \bar{T}_3, \bar{T}_1 + \bar{T}_4, \bar{T}_1 + \bar{T}_2 + \bar{T}_5, t_2 + \bar{T}_6]$, we have a contradiction. This completes the proof of Lemma 4.2•

As in the Simple and Optimistic Models we have shown that the first time π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$ it must return back to this level by a (fixed and finite) duration of most T later, and as before this is easily extended to cover any time that π_1 becomes small. Hence for the Pessimistic Model we also have that if π_1 drops below $\frac{1}{2}\epsilon\pi_1^\star$ at $\tilde{t}_0$ then for $t \in [\tilde{t}_0, \tilde{t}_0 + T]$,

$$\begin{aligned} \frac{d\pi_1}{dt} &\geq -(\mu + \delta_1)\pi_1, \\ \pi_1 &\geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu + \delta_1)(t - \tilde{t}_0)], \\ &\geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu + \delta_1)T], \end{aligned}$$

where T is a fixed duration dependent only on ϵ, Δ and the model parameters. Since $\frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu + \delta_1)T]$ is strictly positive we have that $\pi_{1,\infty} > 0$. Hence we have that (4.24) is true.

4.5.1 Local Stability of Endemic Equilibrium

We now examine local stability of the endemic equilibrium. As in Section 3.7.1 we do not show this directly as this would require examining the roots of a sixth order polynomial, but instead show that a model which is a close approximation to that in

eqns (4.1)-(4.6) (and in particular has the same endemic equilibrium solution) has a locally stable endemic equilibrium when $R_0 > 1$.

As argued previously evidence suggests that the timescale on which addicts inject is of the order of days, whereas that of the other epidemiological and demographic processes is measured in years. Simulations of the Simple and Optimistic Model demonstrated that the behaviour of needles very quickly settles down to a steady relationship with the level of disease among addicts. Moreover this relationship appears to be very close to that at equilibrium. We now use a similar method to that used in Section 3.7.1 to approximate the full Pessimistic Model with an "addict only" model which has the same endemic equilibrium solution but only three rather than six dimensions. Again simulations suggest that this "addict only" model closely mimics the full model.

In the Optimistic Model is was clear that if the prevalences of disease amongst addicts were held at constant values then the prevalences of disease amongst needles would tend to quasi-equilibrium values (with the quasi-equilibrium values being the true endemic equilibrium values if the prevalences of disease amongst addicts were at their endemic equilibrium values). This is also true for the Pessimistic Model. For example treating $\pi_i = \pi_i^+$ for $i = 1, 2, 3$ as constants we have

$$\frac{d\beta_1}{dt} = \lambda\gamma\Big[\pi_1^+ - \beta_1(\pi_1^+ + (1-\pi_1^+)\phi + \hat{\tau})\Big], \tag{4.25}$$

therefore if $\bar{\beta}_1(t) = \pi_1^+/(\pi_1^+ + (1-\pi_1^+)\phi + \hat{\tau})$ is the equilibrium solution to eqn (4.25) then for $\beta_1(t) < \bar{\beta}_1$, $\beta_1(t)$ is increasing in t, and for $\beta_1(t) > \bar{\beta}_1$, $\beta_1(t)$ is decreasing in t. Similarly from the equations

$$\frac{d\beta_{1+3}}{dt} = \lambda\gamma\Big[\pi_{1+3} - \beta_{1+3}(\pi_{1+3} + (1-\pi_{1+3})\phi + \hat{\tau})\Big], \tag{4.26}$$

$$\text{and} \quad \frac{d\beta}{dt} = \lambda\gamma\Big[\pi - \beta(\pi + (1-\pi)\phi + \hat{\tau})\Big], \tag{4.27}$$

we deduce that $\beta_{1+3}(t) \to \bar{\beta}_{1+3}$ and $\beta(t) \to \bar{\beta}$ where $\bar{\beta}_{1+3}$ and $\bar{\beta}$ are obtained by replacing $\pi_1^\star$, $\pi_2^\star$ and $\pi_3^\star$ by π_1^+, π_2^+ and π_3^+ respectively in eqns (4.19) and (4.20). Hence $\beta_i(t) \to \bar{\beta}_i$ for $i = 1, 2, 3$ where $\bar{\beta}_i$ is obtained from eqns (4.13)-(4.15) by replacing $\pi^\star$ by π^+. Therefore intuitively we expect that if disease among addicts spreads much more slowly than among needles then we can approximate the dynamic relationship between $\beta_i(t)$ and $\pi_i(t)$ for $i = 1, 2, 3$ as that obtained from eqn (4.1) by expressing $\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$ in terms of β_1, β_{1+3} and β and then replacing the values of β_1, β_{1+3} and β by the right-hand sides of eqns (4.8), (4.19) and (4.20) respectively with $\pi_i^\star$ replaced by π_i for $i = 1, 2, 3$, together with eqns (4.2) and (4.3).

We can express $\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$ as $\beta_1(\alpha_1 - \alpha_3) + \beta\alpha_2 + \beta_{1+3}(\alpha_3 - \alpha_2)$. Using the equilibrium equations we assume that

$$\beta_1(t) = \frac{\pi_1(t)}{\pi_1(t)(1-\phi) + \hat{\tau} + \phi}, \tag{4.28}$$

$$\beta_{1+3}(t) = \frac{\pi_{1+3}(t)}{\pi_{1+3}(t)(1-\phi) + \hat{\tau} + \phi}, \tag{4.29}$$

$$\text{and } \beta(t) = \frac{\pi(t)}{\pi(t)(1-\phi) + \hat{\tau} + \phi}. \tag{4.30}$$

This gives us the following "addict only" model:

$$\frac{d\pi_1}{dt} = (1-\pi)\lambda(1-\phi)\left[\frac{\pi_1(\alpha_1 - \alpha_3)}{\pi_1(1-\phi) + \hat{\tau} + \phi} + \frac{\pi\alpha_2}{\pi(1-\phi) + \hat{\tau} + \phi} + \frac{\pi_{1+3}(\alpha_3 - \alpha_2)}{\pi_{1+3}(1-\phi) + \hat{\tau} + \phi}\right] - (\mu + \delta_1)\pi_1, \tag{4.31}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu + \delta_2)\pi_2, \tag{4.32}$$

$$\text{and } \frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu + \delta_3)\pi_3. \tag{4.33}$$

We later demonstrate using simulations that as with our previous models this represents a good approximation to our full model.

The Jacobian at $(\pi_1^\star, \pi_2^\star, \pi_3^\star)$ for eqns (4.31)-(4.33) is $\mathbf{J} = (j_{ij})$ where:

$$j_{11} = \frac{\lambda(1-\phi)(\alpha_1 - \alpha_3)}{(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)^2}\left[(1-\pi^\star)(\hat{\tau} + \phi) - \pi_1^{\star 2}(1-\phi) - \pi_1^\star(\hat{\tau} + \phi)\right] - (\mu + \delta_1)$$
$$+ \frac{\lambda(1-\phi)\alpha_2}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2}\left[(1-\pi^\star)(\hat{\tau} + \phi) - \pi^{\star 2}(1-\phi) - \pi^\star(\hat{\tau} + \phi)\right]$$
$$+ \frac{\lambda(1-\phi)(\alpha_3 - \alpha_2)}{(\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi)^2}\left[(1-\pi^\star)(\hat{\tau} + \phi) - \pi_{1+3}^{\star 2}(1-\phi) - \pi_{1+3}^\star(\hat{\tau} + \phi)\right]; \tag{4.34}$$

$$j_{12} = \frac{\lambda(1-\phi)\alpha_2}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2}\left[(1-\pi^\star)(\hat{\tau} + \phi) - \pi^{\star 2}(1-\phi) - \pi^\star(\hat{\tau} + \phi)\right]$$
$$- \frac{\lambda(1-\phi)(\alpha_1 - \alpha_3)\pi_1^\star}{\pi_1^\star(1-\phi) + \hat{\tau} + \phi} - \frac{\lambda(1-\phi)(\alpha_3 - \alpha_2)\pi_{1+3}^\star}{\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi}; \tag{4.35}$$

$$j_{13} = \frac{\lambda(1-\phi)(\alpha_3 - \alpha_2)}{(\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi)^2}\left[(1-\pi^\star)(\hat{\tau} + \phi) - \pi_{1+3}^{\star 2}(1-\phi) - \pi_{1+3}^\star(\hat{\tau} + \phi)\right]$$
$$+ \frac{\lambda(1-\phi)\alpha_2}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2}\left[(1-\pi^\star)(\hat{\tau} + \phi) - \pi^{\star 2}(1-\phi) - \pi^\star(\hat{\tau} + \phi)\right]$$
$$- \frac{\lambda(1-\phi)(\alpha_1 - \alpha_3)\pi_1^\star}{\pi_1^\star(1-\phi) + \hat{\tau} + \phi}; \tag{4.36}$$

$j_{21} = \delta_1$; $j_{22} = -(\mu + \delta_2)$; $j_{23} = 0$; $j_{31} = 0$; $j_{32} = \delta_2$ and $j_{33} = -(\mu + \delta_3)$. We require to show that $a_1, a_2, a_3 > 0$ and $a_1 a_2 > a_3$ where a_1, a_2 and a_3 are as defined in eqns (3.27)-(3.29). First note that

$$\lambda(1-\phi)(1-\pi^\star)$$

$$\times \left[\frac{\alpha_1 - \alpha_3}{(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)^2} + \frac{\alpha_2}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2} + \frac{\alpha_3 - \alpha_2}{(\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi)^2} \right],$$

$$= \lambda(1-\phi)(1-\pi^\star)$$

$$\times \left[\frac{(\alpha_1 - \alpha_3)\beta_1^\star}{\pi_1^\star(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)} + \frac{\alpha_2 \beta^\star}{\pi^\star(\pi^\star(1-\phi) + \hat{\tau} + \phi)} + \frac{(\alpha_3 - \alpha_2)\beta_{1+3}^\star}{\pi_{1+3}^\star(\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi)} \right],$$

$$\leq \frac{\lambda(1-\phi)(1-\pi^\star)}{\pi_1^\star(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)} \left[(\alpha_1 - \alpha_3)\beta_1^\star + \alpha_2 \beta^\star + (\alpha_3 - \alpha_2)\beta_{1+3}^\star \right],$$

$$= \frac{(\mu + \delta_1)(\hat{\tau} + \phi)}{\pi_1^\star(1-\phi) + \hat{\tau} + \phi}.$$

Using the eqns (4.34)-(4.36) and (3.27)-(3.29) we find that

$$a_1 \geq \mu + \delta_1 + \mu + \delta_2 + \mu + \delta_3 - \left\{ \lambda(1-\phi)(1-\pi^\star)(\hat{\tau} + \phi) \right.$$

$$\left. \times \left[\frac{\alpha_1 - \alpha_3}{(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)^2} + \frac{\alpha_2}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2} + \frac{\alpha_3 - \alpha_2}{(\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi)^2} \right] \right\},$$

$$\geq \mu + \delta_1 + \mu + \delta_2 + \mu + \delta_3 - \frac{(\mu + \delta_1)(\hat{\tau} + \phi)}{\pi_1^\star(1-\phi) + \hat{\tau} + \phi} > 0.$$

In a similar fashion we have that

$$a_2 \geq (\mu + \delta_2)(\mu + \delta_3) + (\mu + \delta_1)(\mu + \delta_2) + (\mu + \delta_1)(\mu + \delta_3)$$

$$- \frac{\delta_1 \lambda(1-\phi)\alpha_2(1-\pi^\star)(\hat{\tau} + \phi)}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2} - \left\{ (\mu + \delta_2 + \mu + \delta_3)\lambda(1-\phi)(1-\pi^\star)(\hat{\tau} + \phi) \right.$$

$$\left. \times \left[\frac{\alpha_1 - \alpha_3}{(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)^2} + \frac{\alpha_2}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)^2} + \frac{\alpha_3 - \alpha_2}{(\pi_{1+3}^\star(1-\phi) + \hat{\tau} + \phi)^2} \right] \right\},$$

$$\geq (\mu + \delta_2)(\mu + \delta_3) + (\mu + \delta_1)(\mu + \delta_2) + (\mu + \delta_1)(\mu + \delta_3)$$

$$- (\mu + \delta_2 + \mu + \delta_3)(\mu + \delta_1) \frac{(\hat{\tau} + \phi)}{\pi_1^\star(1-\phi) + \hat{\tau} + \phi}$$

$$+ \frac{(\mu + \delta_2 + \mu + \delta_3)\lambda(1-\phi)(1-\pi^\star)(\hat{\tau} + \phi)\alpha_2 \pi^\star}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)} \left(\frac{1}{\pi_1^\star} - \frac{1}{\pi^\star} \right)$$

$$- \frac{\delta_1 \lambda(1-\phi)\alpha_2(1-\pi^\star)(\hat{\tau} + \phi)}{(\pi^\star(1-\phi) + \hat{\tau} + \phi)(\pi_1^\star(1-\phi) + \hat{\tau} + \phi)},$$

$$= (\mu+\delta_2)(\mu+\delta_3)+\frac{(\mu+\delta_2+\mu+\delta_3)(\mu+\delta_1)\pi_1^\star(1-\phi)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi}$$

$$+\frac{\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)\alpha_2}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)}\Big\{(L-1)(\mu+\delta_2+\mu+\delta_3)-\delta_1\Big\}>0.$$

We have that

$$a_3 \geq (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)-\Bigg\{(\mu+\delta_2)(\mu+\delta_3)\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)$$

$$\times\left[\frac{\alpha_1-\alpha_3}{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)^2}+\frac{\alpha_2}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}+\frac{\alpha_3-\alpha_2}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2}\right]\Bigg\}$$

$$-\frac{\delta_1(\mu+\delta_3)\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)\alpha_2}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}$$

$$-\Bigg\{\delta_1\delta_2\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)$$

$$\times\left[\frac{\alpha_2}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}+\frac{\alpha_3-\alpha_2}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2}\right]\Bigg\},$$

$$\geq (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)\left(1-\frac{\hat{\tau}+\phi}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi}\right)$$

$$+\frac{(\mu+\delta_2)(\mu+\delta_3)\lambda(1-\pi^\star)(1-\phi)(\hat{\tau}+\phi)\alpha_2\pi^\star}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)}\left(\frac{1}{\pi_1^\star}-\frac{1}{\pi^\star}\right)$$

$$+\frac{(\mu+\delta_2)(\mu+\delta_3)\lambda(1-\pi^\star)(1-\phi)(\hat{\tau}+\phi)(\alpha_3-\alpha_2)\pi_{1+3}^\star}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)}\left(\frac{1}{\pi_1^\star}-\frac{1}{\pi_{1+3}^\star}\right)$$

$$-\frac{\delta_1(\mu+\delta_3)\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)\alpha_2}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)}$$

$$-\left\{\frac{\delta_1\delta_2\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi}\left[\frac{\alpha_2}{\pi^\star(1-\phi)+\hat{\tau}+\phi}+\frac{\alpha_3-\alpha_2}{\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi}\right]\right\},$$

$$= (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)\frac{\pi_1^\star(1-\phi)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi}>0.$$

We now wish to show that $a_1a_2-a_3>0$. Using eqns (3.27)-(3.29) we have that

$$a_1a_2-a_3 = \Big[(\mu+\delta_2+\mu+\delta_3)(\mu+\delta_3-j_{11})-\delta_1j_{12}\Big](\mu-j_{11})$$

$$+\delta_2(\mu+\delta_2+\mu+\delta_3)(\mu+\delta_3-j_{11})+\delta_1\delta_2(j_{13}-j_{12}). \quad (4.37)$$

From eqn (4.37) it is sufficient to show that (i) $j_{11}<0$, (ii) $j_{13}>j_{12}$ and (iii) $(\mu+\delta_2+\mu+\delta_3)(\mu+\delta_3-j_{11})>\delta_1j_{12}$. We first show (i):

$$-j_{11} > \mu+\delta_1-\Big\{\lambda(1-\phi)(1-\pi^\star)(\hat{\tau}+\phi)$$

$$\times \left[\frac{\alpha_1 - \alpha_3}{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)^2} + \frac{\alpha_2}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2} + \frac{\alpha_3 - \alpha_2}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2} \right] \Bigg\},$$

$$> \quad \mu + \delta_1 - \frac{(\hat{\tau}+\phi)(\mu+\delta_1)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi},$$

$$= \quad \frac{(\mu+\delta_1)(1-\phi)\pi_1^\star}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi} > 0.$$

By examining the terms j_{13} and j_{12} it is obvious that $j_{13} > j_{12}$. It is also easy to see that $(\mu+\delta_2+\mu+\delta_3)(\mu+\delta_3-j_{11}) - \delta_1 j_{12} \geq a_2 > 0$. Hence (ii) and (iii) follow, $a_1 a_2 > a_3$ and all the Routh-Hurwitz conditions are satisfied when $R_0 > 1$. Therefore we have shown that the "addict only" approximation to the full Pessimistic Model has a locally stable endemic equilibrium when $R_0 > 1$. As in the Optimistic Model it is possible to make this approximation argument more rigorous by showing that if $\lambda\gamma$ is large compared with the other parameters of the model apart from τ, (including $\lambda\alpha_1$, $\lambda\alpha_2$ and $\lambda\alpha_3$) then three of the roots of the characteristic equation of the Jacobian of the full model at the endemic equilibrium have strictly negative real parts and the other three are close to the roots of the characteristic equation of the Jacobian of the "addict-only" model at the endemic equilibrium. See Appendix D for details.

4.5.2 Sufficient Conditions for Global Stability of $(\pi^\star, \beta^\star)$

We now derive sufficient conditions for the disease to tend to the endemic equilibrium if it is initially present in the Pessimistic Model using a similar method to that in Section 3.7.2. As in the previous models we do not use the coordinate system $(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$ directly but the translated form where the origin of the new system corresponds to the endemic equilibrium, $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star)$, in the original form. This translation gives the following set of model equations:

$$\frac{d\tilde{\pi}_1}{dt} = (1-\pi^\star)\lambda(1-\phi)(\tilde{\beta}_1\alpha_1 + \tilde{\beta}_2\alpha_2 + \tilde{\beta}_3\alpha_3) - \tilde{\pi}\lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)$$

$$-(\mu+\delta_1)\tilde{\pi}_1, \tag{4.38}$$

$$\frac{d\tilde{\pi}_2}{dt} = \delta_1\tilde{\pi}_1 - (\mu+\delta_2)\tilde{\pi}_2, \tag{4.39}$$

$$\frac{d\tilde{\pi}_3}{dt} = \delta_2\tilde{\pi}_2 - (\mu+\delta_3)\tilde{\pi}_3, \tag{4.40}$$

$$\frac{d\tilde{\beta}_1}{dt} = \lambda\gamma(1-\beta_1^\star(1-\phi))\tilde{\pi}_1 - (\lambda\gamma\phi + \lambda\gamma\pi_1(1-\phi)+\tau)\tilde{\beta}_1, \tag{4.41}$$

$$\frac{d\tilde{\beta}_2}{dt} = -\lambda\gamma\beta_2^\star(1-\phi)\tilde{\pi}_1 + \lambda\gamma(1-\beta^\star(1-\phi))\tilde{\pi}_2 - \lambda\gamma\beta_2^\star(1-\phi)\tilde{\pi}_3 - \lambda\gamma\pi_2(1-\phi)\tilde{\beta}_1$$

$$-\lambda\gamma(\phi+\hat{\tau}+\pi(1-\phi))\tilde{\beta}_2 - \lambda\gamma(1-\phi)\pi_2\tilde{\beta}_3, \tag{4.42}$$

and $$\frac{d\tilde{\beta}_3}{dt} = -\lambda\gamma(1-\phi)\beta_3^\star\tilde{\pi}_1 + \lambda\gamma(1-\beta_1^\star(1-\phi)-\beta_3^\star(1-\phi))\tilde{\pi}_3 - \lambda\gamma(1-\phi)\pi_3\tilde{\beta}_1$$

$$-\lambda\gamma(\phi+\hat{\tau}+\pi_1(1-\phi)+\pi_3(1-\phi))\tilde{\beta}_3, \tag{4.43}$$

where $\tilde{\beta}_i = \beta_i - \beta_i^\star$, $\tilde{\pi}_i = \pi_i - \pi_i^\star$ for $i = 1,2,3$, and $\tilde{\pi} = \tilde{\pi}_1 + \tilde{\pi}_2 + \tilde{\pi}_3$. The system represented by eqns (4.38)-(4.43) can be represented in vector form as

$$\frac{d\tilde{\mathbf{x}}}{dt} = \mathbf{V}(\mathbf{x})\tilde{\mathbf{x}},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$, $\tilde{\mathbf{x}}^T = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3)$ and $\mathbf{V}(\mathbf{x}) =$

$$\begin{bmatrix}
-(\mu+\delta_1)-\lambda\hat{\beta}(1-\phi) & -\lambda\hat{\beta}(1-\phi) & -\lambda\hat{\beta}(1-\phi) & (1-\pi^*)\lambda\alpha_1(1-\phi) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 \\
\lambda\gamma(1-\beta_1^\star(1-\phi)) & 0 & 0 & -\lambda\gamma(\phi+\hat{\tau}+\pi_1(1-\phi)) \\
-\lambda\gamma\beta_2^\star(1-\phi) & \lambda\gamma(1-\beta^*(1-\phi)) & -\lambda\gamma\beta_2^\star(1-\phi) & -\lambda\gamma\pi_2(1-\phi) \\
-\lambda\gamma\beta_3^\star(1-\phi) & 0 & \lambda\gamma(1-\beta_{1+3}^\star(1-\phi)) & -\lambda\gamma\pi_3(1-\phi)
\end{bmatrix}$$

$$\begin{bmatrix}
(1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
-\lambda\gamma(\phi+\hat{\tau}+\pi(1-\phi)) & -\lambda\gamma\pi_2(1-\phi) \\
0 & -\lambda\gamma(\phi+\hat{\tau}+\pi_{1+3}(1-\phi))
\end{bmatrix},$$

where $\hat{\beta} = \beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3$, $\beta_{1+3}^\star = \beta_1^\star + \beta_3^\star$ and $\pi_{1+3} = \pi_1 + \pi_3$. Now consider $\tilde{\mathbf{y}}^T = (\tilde{\pi}, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}, \tilde{\beta}_{1+3})$. We can write

$$\frac{d\tilde{\mathbf{y}}}{dt} = \mathbf{W}(\mathbf{x})\tilde{\mathbf{y}},$$

where $\tilde{\mathbf{y}}^T = (\tilde{\pi}, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}, \tilde{\beta}_{1+3})$ and $\mathbf{W}(\mathbf{x}) =$

$$\begin{bmatrix}
-\mu-\lambda\hat{\beta}(1-\phi) & 0 & -\delta_3 & (1-\pi^*)\lambda(\alpha_1-\alpha_3)(1-\phi) \\
\delta_1 & -(\mu+\delta_2+\delta_1) & -\delta_1 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 \\
\lambda\gamma(1-\beta_1^\star(1-\phi)) & -\lambda\gamma(1-\beta_1^\star(1-\phi)) & -\lambda\gamma(1-\beta_1^\star(1-\phi)) & -\lambda\gamma(\phi+\hat{\tau}+\pi_1(1-\phi)) \\
(1-\beta^*(1-\phi))\lambda\gamma & 0 & 0 & 0 \\
\lambda\gamma(1-\beta_{1+3}^\star(1-\phi)) & -\lambda\gamma(1-\beta_{1+3}^\star(1-\phi)) & 0 & 0
\end{bmatrix}$$

$$\begin{matrix} (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda(\alpha_3-\alpha_2)(1-\phi) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -\lambda\gamma(\phi+\hat{\tau}+\pi(1-\phi)) & 0 \\ 0 & -\lambda\gamma(\phi+\hat{\tau}+\pi_{1+3}(1-\phi)) \end{matrix}\Bigg].$$

We have that $\tilde{\mathbf{y}} = \mathbf{J}_2\tilde{\mathbf{x}}$ where

$$\mathbf{J}_2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix},$$

and that $\mathbf{W}(\mathbf{x})$ contains elements of $\mathbf{x}$ only on the leading diagonal. Using Theorem 4.4 we can replace the variables $\hat{\beta}$, π_1, π_{1+3} and π in $\mathbf{W}(\mathbf{x})$ with a constant lower bound, $\bar{\varepsilon}$. Hence we have that for $t \geq T^+$, $\mathbf{W}(\mathbf{x}) \leq \mathbf{W}^+ = \mathbf{W}(\mathbf{0}) - \bar{\varepsilon}\mathbf{E}$ where $\mathbf{E} =$

$$\begin{bmatrix} \lambda(1-\phi) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda\gamma(1-\phi) & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda\gamma(1-\phi) & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda\gamma(1-\phi) \end{bmatrix}.$$

The form of $\mathbf{W}(\mathbf{x})$ and $\mathbf{W}^+$ are similar to those discussed in Section 2.5.1 and therefore we again require that $\mathbf{W}^+$ is Volterra-Lyapunov stable for $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star)$ to be globally stable when $R_0 > 1$. (We can show that $\mathbf{W}^+$ is Lyapunov stable in a similar fashion to our previous models however the proof of this is rather long and can be found in Appendix D). By considering the matrix $\mathbf{S} = -\mathbf{W}^{+T}$ we find that $\mathcal{M}(\mathbf{S}) =$

$$\begin{bmatrix} \mu+\lambda\varepsilon(1-\phi) & -\delta_1 & 0 & -\lambda\gamma(1-\beta_1^*(1-\phi)) & -\lambda\gamma(1-\beta^*(1-\phi)) & -\lambda\gamma(1-\beta_{1+3}^*(1-\phi)) \\ 0 & \mu+\delta_1+\delta_2 & -\delta_2 & -\lambda\gamma(1-\beta_1^*(1-\phi)) & 0 & -\lambda\gamma(1-\beta_{1+3}^*(1-\phi)) \\ -\delta_3 & -\delta_1 & \mu+\delta_3 & -\lambda\gamma(1-\beta_1^*(1-\phi)) & 0 & 0 \\ -(1-\pi^*)\lambda(\alpha_1-\alpha_3)(1-\phi) & 0 & 0 & \lambda\gamma(\phi+\hat{\tau}+\varepsilon(1-\phi)) & 0 & 0 \\ -(1-\pi^*)\lambda\alpha_2(1-\phi) & 0 & 0 & 0 & \lambda\gamma(\phi+\hat{\tau}+\varepsilon(1-\phi)) & 0 \\ -(1-\pi^*)\lambda(\alpha_3-\alpha_2)(1-\phi) & 0 & 0 & 0 & 0 & \lambda\gamma(\phi+\hat{\tau}+\varepsilon(1-\phi)) \end{bmatrix}.$$

As with our previous models a sufficient condition for Volterra-Lyapunov stability is that all the principal minors of $\mathcal{M}(\mathbf{S})$ are strictly positive. Again by examining each of the individual principal minors it is straightforward to verify that if $(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2>0$ then $\det(\mathcal{M}(\mathbf{S}))>0$ is a necessary and sufficient condition for the positivity of all principal minors, where $\det(\mathcal{M}(\mathbf{S}))=$

$$-(1-\pi^\star)\lambda(\alpha_3-\alpha_2)(1-\phi)\lambda^3\gamma^3(\phi+\hat{\tau}+\bar{\varepsilon}(1-\phi))^2\delta_1(\mu+\delta_3)(1-\beta_{1+3}(1-\phi))$$

$$-\delta_1\lambda^3\gamma^3(\phi+\hat{\tau}+\bar{\varepsilon}(1-\phi))^2(1-\pi^\star)\lambda(\alpha_1-\alpha_3)(1-\phi)[\delta_2+\mu+\delta_3](1-\beta_1^\star(1-\phi))$$

$$-\lambda^3\gamma^3(\phi+\hat{\tau}+\bar{\varepsilon}(1-\phi))^3\delta_1\delta_2\delta_3$$

$$-\Big[(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2\Big](1-\pi^\star)\lambda(1-\phi)\lambda^3\gamma^3(\phi+\hat{\tau}+\bar{\varepsilon}(1-\phi))^2$$

$$\times\Big\{(\alpha_3-\alpha_2)(1-\beta_{1+3}(1-\phi))+\alpha_2(1-\beta^\star(1-\phi))+(\alpha_1-\alpha_3)(1-\beta_1^\star(1-\phi))\Big\}$$

$$+\lambda^3\gamma^3(\phi+\hat{\tau}+\bar{\varepsilon}(1-\phi))^3(\mu+\lambda\bar{\varepsilon}(1-\phi))\Big[(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2\Big].$$

Therefore if these conditions are satisfied, $R_0>1$ and disease is initially present then the prevalence of disease in our model eventually tends to the endemic equilibrium solution. As before $\det(\mathcal{M}(\mathbf{S}))>0$ is a genuine condition, but one which is not always true. For example if $\mu\gg\delta_1,\delta_2,\delta_3,\lambda\alpha_1,\lambda\alpha_2,\lambda\alpha_3$ then it will be true. However for $\bar{\varepsilon}$ near zero and $\alpha_1>0$ but $\alpha_2=\alpha_3=0$ then if $(\mu+\delta_1+\delta_2)(\mu+\delta_3)-\delta_1\delta_2>0$ we have that $\det(\mathcal{M}(\mathbf{S}))>0$ implies that

$$\begin{aligned}\mu(\phi+\hat{\tau}) &> \lambda\alpha_1(1-\phi)(1-\beta_1^\star(1-\phi))(1-\pi^\star),\\ &= \frac{\lambda\alpha_1(1-\phi)(\hat{\tau}+\phi)(1-\pi^\star)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi},\end{aligned}$$

so

$$\mu > \frac{\lambda\alpha_1(1-\phi)(1-\pi^\star)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi}.$$

However using the equilibrium equations we have that

$$(\mu+\delta_1)\pi_1^\star = \frac{\lambda\alpha_1(1-\phi)(1-\pi^\star)\pi_1^\star}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi},$$

so

$$\mu < \frac{\lambda\alpha_1(1-\phi)(1-\pi^\star)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi},$$

and therefore $\det(\mathcal{M}(\mathbf{S}))<0$.

Having examined the behaviour of the Pessimistic Model analytically we now use simulation to examine the dynamic behaviour of our model. Also of interest is whether

$R_0 > 1$ on its own is a necessary and sufficient condition for the disease to tend to the unique endemic equilibrium if it is initially present.

4.6 Simulation Study of the Pessimistic Model

In order to validate the theoretical results of the previous section we now demonstrate the behaviour of the Pessimistic Model graphically using numerical simulations. We adopt a very similar approach as for the Optimistic Model and use the same parameter estimates in our simulations. We wish to demonstrate two key properties of the Pessimistic Model. If the parameter estimates are such that $R_0 > 1$ then provided that the disease is present in at least one addict or one needle then it spreads among the population until a steady state is reached where a fraction $\pi^\star$ of all addicts are infected with proportions $\pi_1^\star$, $\pi_2^\star$ and $\pi_3^\star$ in the three infective stages and a fraction $\beta^\star$ of all needles are infected with proportions $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ in the three infective stages. We also wish to demonstrate that if the parameter estimates give rise to a value of $R_0 \leq 1$ then the disease will die out in both addicts and needles.

We use two sets of parameters to illustrate the properties of the Pessimistic Model. The first set of parameters uses the estimates from Table 3.1 and gives a value for R_0 of 3.317 (this is not the same as in the Optimistic Model since the expression of R_0 for the Pessimistic Model is different). The second set of parameters is the same as the first with the exception of ϕ (the probability that an addict successfully cleans a needle prior to use) which is now 0.87, this gives a value for R_0 of approximately 0.908.

Figure 4.1 shows the Pessimistic Model simulated over forty years using the set of parameter estimates where $R_0 = 3.317$. At time zero we have assumed that one percent of the total population of addicts are in stage one infectivity, at this time no other addicts or needles are infectious. The figure shows the progress of each type of infectious addict and each type of infectious needle over time. It is clear that the fraction of infected addicts in each stage eventually reaches a steady state as does the fraction of infected needles in each state. The steady state values in these simulations are $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star) = (0.028, 0.355, 0.249, 0.039, 0.333, 0.304)$, which correspond to $\pi^\star = 0.633$ and $\beta^\star = 0.675$.

Figures 4.2 and 4.3 show simulations of the total prevalence of disease in addicts and the total prevalence of disease in needles in the Pessimistic Model for four different sets of initial conditions and where $R_0 = 3.317$. In each initial condition we assumed

Figure 4.1: Pessimistic Model when $R_0 > 1$

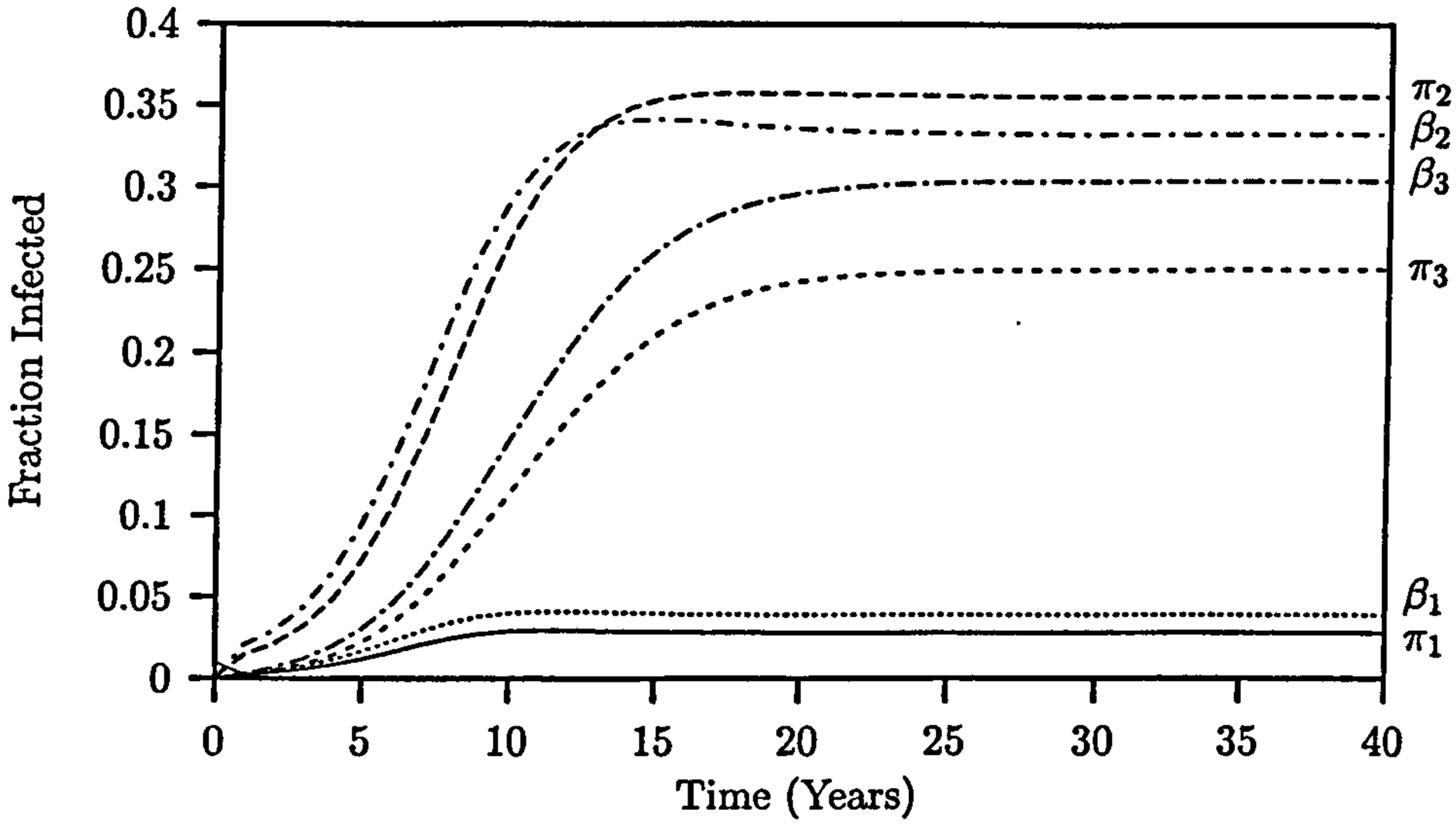


Figure 4.2: Pessimistic Model when $R_0 > 1$

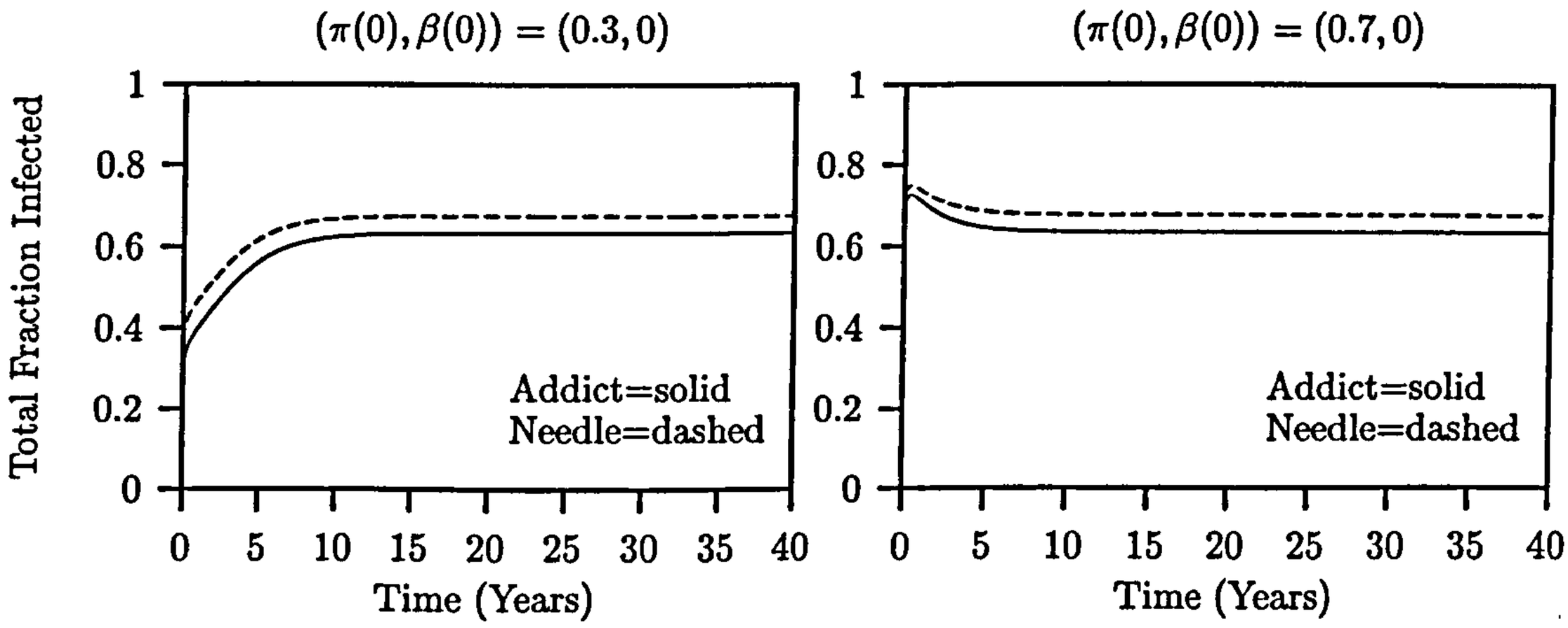
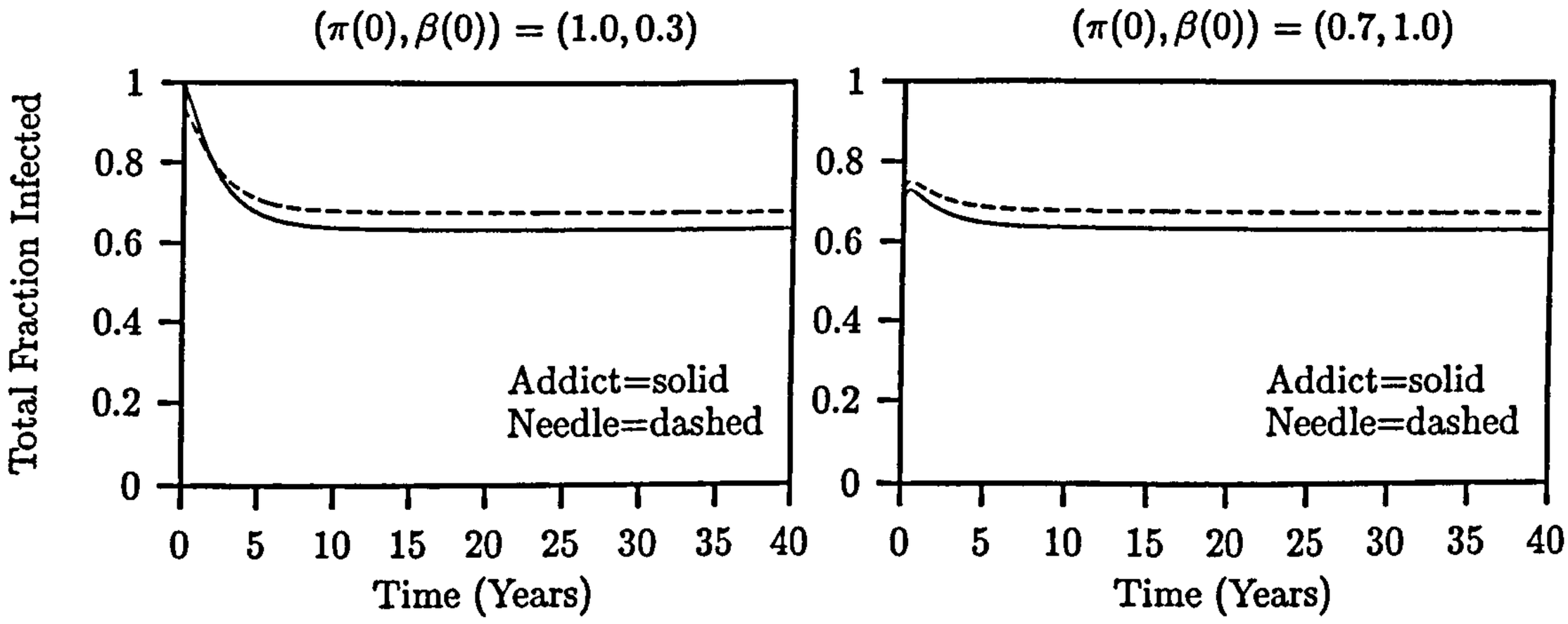


Figure 4.3: Pessimistic Model when $R_0 > 1$



that $\pi(0) = \pi_1(0)$ and $\beta(0) = \beta_1(0)$ with all other components starting at zero. It is clear that in each set of simulations the model tends to the same endemic equilibrium solution (which is also the same as that shown in Figure 4.1). Other simulations (not illustrated) suggest that for a wide variety of parameter estimates and initial conditions provided that disease is initially present and $R_0 > 1$ then the system will approach the unique endemic equilibrium.

An interesting feature of the simulations in Figures 4.2 and 4.3 is the very fast initial movement of $\beta(t)$ compared to $\pi(t)$. We have previously argued that if this were the case then our "addict only" approximation discussed in Section 4.5.1 may be appropriate. It appears that the prevalences of disease in needles very quickly settle down to quasi-equilibrium values which are determined by the prevalences of disease in addicts using the same relationships as in eqns (4.8), (4.19) and (4.20) between the equilibrium prevalences of disease in needles and the equilibrium prevalences of disease in addicts. Indeed as in the Optimistic Model we find that if we start with comparable initial conditions then the dynamic behaviour between the full model and the corresponding "addict only" approximation is indistinguishable.

We now simulate the Pessimistic Model using the same set of parameter estimates as previously except that now $\phi = 0.87$ which reduces R_0 to 0.908. Figure 4.4 shows the Pessimistic Model simulated over 160 years. At time zero we have assumed that the population is in an endemic steady state where 2.8% of the total population of addicts are in stage one infectivity, 35.5% are in stage two infectivity and 24.9% are in stage three infectivity. We also assume that 3.9% of the total population of needles are in state one infectivity, 33.2% are in state two infectivity and 30.4% are in state three infectivity. These values correspond to the endemic equilibrium for the Pessimistic Model using the first set of parameters. We now suppose that at time zero R_0 has been reduced from 3.317 down to 0.908. As in Figure 4.1 each line on the figure represents the spread of the various stages of infectivity among the addict population and needle population. It is clear from the figure that the disease dies out in all addicts and all needles and after about 150 years the Pessimistic Model reaches the disease-free equilibrium.

Figures 4.5 and 4.6 show the behaviour of the total prevalence of disease among addicts in our model for the four sets of initial conditions: $\pi_1(0) = 0.3$ and $\beta_1(0) = 0.0$; $\pi_1(0) = 0.7$ and $\beta_1(0) = 0.0$; $\pi_1(0) = 1.0$ and $\beta_1(0) = 0.3$; and $\pi_1(0) = 0.7$ and $\beta_1(0) = 1.0$. In each of these cases no other types of addicts or needles are initially

Figure 4.4: System Tends to Disease-Free Equilibrium when $R_0 < 1$

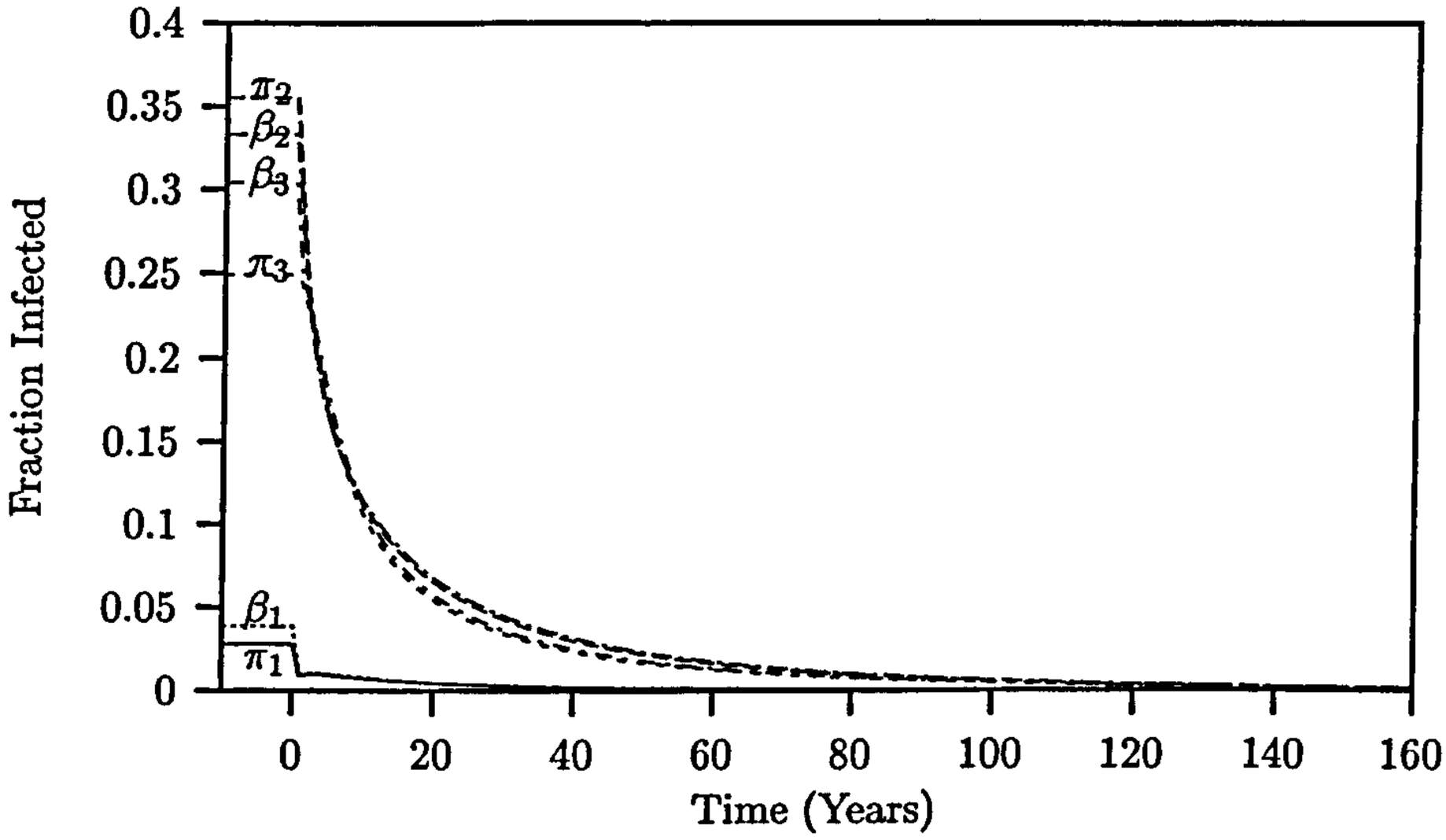


Figure 4.5: Pessimistic Model when $R_0 < 1$

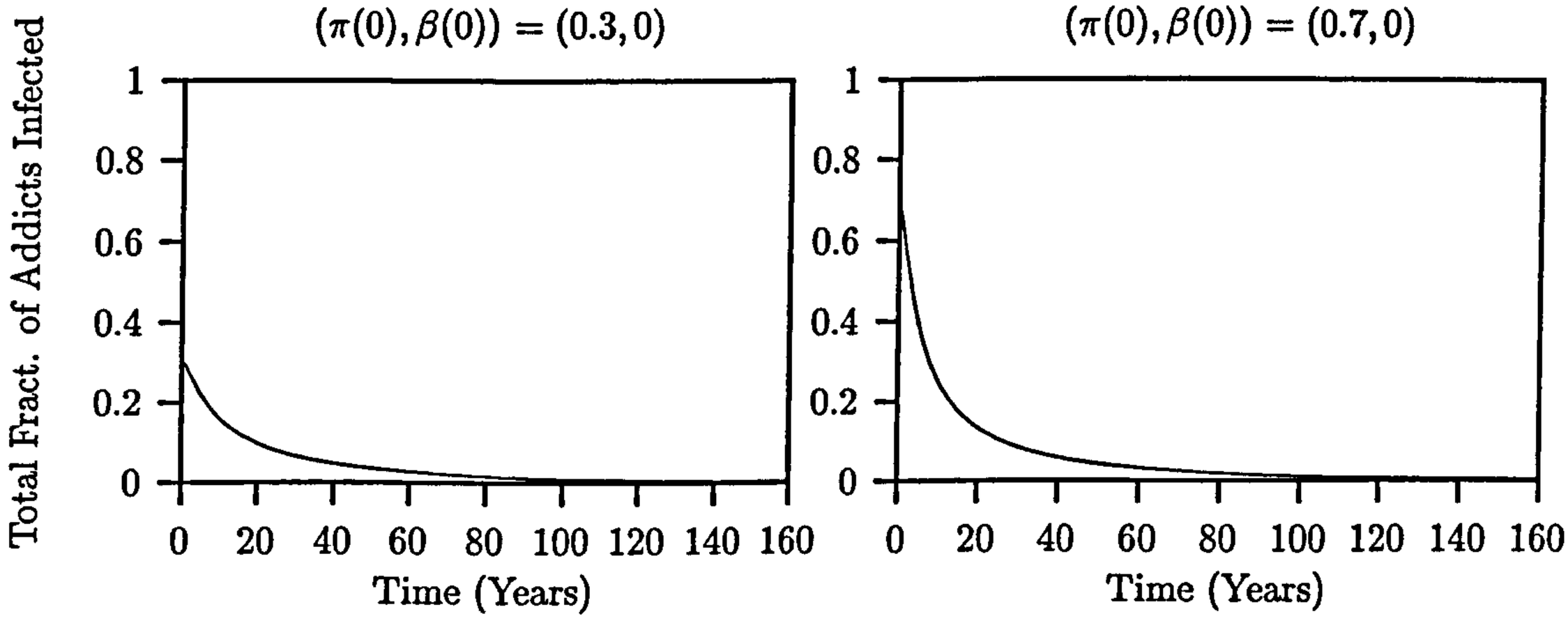
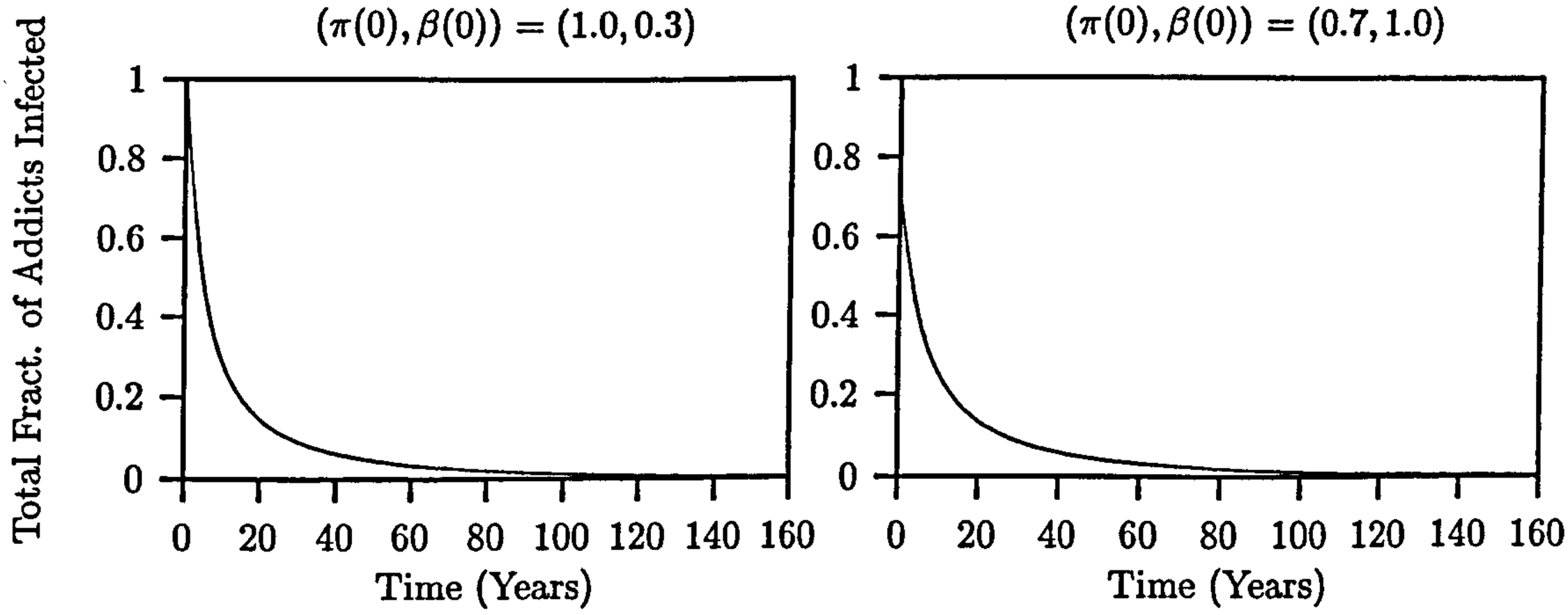


Figure 4.6: Pessimistic Model when $R_0 < 1$



infectious and $R_0 = 0.908$. We have not illustrated the total prevalence of disease among needles but this behaves very similarly. It is clear that eventually the model reaches the disease-free equilibrium solution. Many simulations of this model were carried out for a variety of parameter estimates and initial conditions. In each case we found that if $R_0 < 1$ then disease will be eventually eradicated from the population. This is consistent with Theorem 4.2.

4.7 Summary of Results for the Pessimistic Model

We began the chapter by a discussing a set of addict-needle interaction assumptions which were more pessimistic than would reasonably be expected to occur in reality. By pessimistic we mean that under these assumptions we would expect the disease to travel faster, reach a higher long term prevalence level and be less responsive to control measures than would realistically be the case. We then derived a system of differential equations which contained these pessimistic assumptions and computed the basic reproductive number for this model.

Once R_0 had been computed we then moved on to deriving analytical results relating to the stability of the equilibrium solutions of our model. We showed that if $R_0 \leq 1$ then the disease-free equilibrium is globally stable, and if $R_0 > 1$ and disease is initially present then it will persist among the population for all time. We then showed that a simplified version of the Pessimistic Model has an endemic equilibrium solution which is locally stable when $R_0 > 1$. Moreover we showed that if $\lambda\gamma$ is much larger than the other model parameters apart from τ, the needle exchange rate, (which we expect to be the case for realistic parameter estimates) then our full model is also locally stable when $R_0 > 1$. We then derived sufficient conditions for the prevalence of disease to tend to the unique endemic equilibrium solution. We finally carried out a small number of simulations of our model in order to verify these results and determine whether $R_0 > 1$ is a necessary and sufficient condition for the disease to approach the unique endemic equilibrium provided that it is initially present.

In the previous chapter we investigated our best case scenario model, the Optimistic Model, in this chapter we examined our worst case scenario model, the Pessimistic Model. We have shown that these two models have qualitatively very similar theoretical properties, and it now remains for us to compare and contrast these models with the Simple Model and the original Kaplan and O'Keefe Model. However before we draw

conclusions as to the potential effects of three stage infectivity we first examine an additional three stage model, the General Model. So far our investigation into three stage infectivity in both addicts and needles has relied on models which have had fixed assumptions relating to how addicts and needles interact with each other. In the next chapter we examine a model which is not limited to a single set of addict-needle interactions assumptions but can incorporate a wide variety of different assumptions.

Chapter 5

The General Mixing Model

5.1 Introduction

We have previously discussed the Optimistic and Pessimistic Models. These models include three types of infectious addicts and three types of infectious needles but assume that addicts and needles interact in very specific ways. The Optimistic Model assumes that a needle is always left in the same infectious state as the last user while the Pessimistic Model assumes that a needle is always left in the more infectious state between that of the current user and that of the needle prior to use. In this chapter we develop and investigate a model which has a more general addict-needle interaction structure than the Optimistic and Pessimistic Models and where the Optimistic and Pessimistic Models are special cases of this more general model.

The fundamental difference between our previous models and the model discussed in this chapter is the use of a probability structure to define the outcomes of each addict-needle interaction. In the Optimistic and Pessimistic Models the outcome of each addict-needle interaction occurred with either probability one or probability zero. For example consider the event where an addict in stage two infectivity injects with a needle in state one infectivity. In the Optimistic Model this needle would be left in state two infectivity (with probability one) whereas in the Pessimistic Model the needle would remain in state one infectivity (with probability one). In contrast, in the model in this chapter a needle in state one infectivity (after use by an addict in stage two infectivity) can either remain in state one infectivity or enter state two or state three infectivity where each of these outcomes occurs with a given probability and where these three probabilities sum to one.

We refer to the model in this chapter as the General Model on account of the wide range of addict-needle interaction assumptions which it can incorporate. We investigate the General Model in a similar fashion to the Simple, Optimistic and Pessimistic models. We first derive the differential equations which define our model and then compute the basic reproductive number. Next we move on to deriving analytical results, we are particularly interested in whether the prevalence of disease tends to an endemic equilibrium if $R_0 > 1$ and disease is initially present and tends to the disease-free equilibrium if $R_0 \leq 1$. We then conduct a simulation study to investigate the behaviour of the General Model for a variety of addict-needle interaction assumptions. The chapter concludes with a summary of the main points.

5.2 Addict-Needle Interaction Structure

In this chapter we wish to investigate the effect of different addict-needle interaction assumptions on the spread of HIV via needle sharing. In order to do this sensibly we need to use a model which can incorporate a wide range of different addict-needle interaction assumptions. To achieve this we assign a probability to the outcome of each addict-needle interaction. In this way we can construct a model by considering the outcome of each addict-needle interaction and the probability that this outcome occurs. Let $p^{\star}_{ijk}$ denote the probability that a needle in state i infectivity will be left in state k infectivity immediately after use by an addict in stage j infectivity. We now outline all the different events and outcomes which can arise between addicts and needles. Let 0, 1, 2 and 3 denote the infectious states: Uncontaminated, Acute Infectivity, Asymptomatic and Pre-AIDS Symptoms respectively. Hence there are a total of 16 possible addict-needle interactions each with four possible outcomes where each outcome occurs with probability $p^{\star}_{ijk}$. Table 5.1 illustrates each of these events.

Ascribing a precise numerical estimate to many of the $p^{\star}_{ijk}$ terms in Table 5.1 is at best difficult, however we can argue that a number of these terms should be set to either zero or one. Common sense dictates that a state i infectious needle should remain a state i infectious needle after use by an addict in stage i infectivity, this implies that $p^{\star}_{000} = p^{\star}_{111} = p^{\star}_{222} = p^{\star}_{333} = 1$. A further common sense criteria is that the viral load of a needle after use must be less than or equal to the maximum of the viral load of the current user and the state of the needle prior to use. For example a needle in state two infectivity used by an addict in stage three infectivity cannot be left in state one

Table 5.1: Addict-Needle Interactions

Addict Infectivity	Needle Infectivity (prior to use)			
	Uninfectious	Acute	Asymptomatic	Pre-AIDS
Uninfectious Addicts	$p^\star_{000}$ $p^\star_{001}$ $p^\star_{002}$ $p^\star_{003}$	$p^\star_{100}$ $p^\star_{101}$ $p^\star_{102}$ $p^\star_{103}$	$p^\star_{200}$ $p^\star_{201}$ $p^\star_{202}$ $p^\star_{203}$	$p^\star_{300}$ $p^\star_{301}$ $p^\star_{302}$ $p^\star_{303}$
Acute Addicts	$p^\star_{010}$ $p^\star_{011}$ $p^\star_{012}$ $p^\star_{013}$	$p^\star_{110}$ $p^\star_{111}$ $p^\star_{112}$ $p^\star_{113}$	$p^\star_{210}$ $p^\star_{211}$ $p^\star_{212}$ $p^\star_{213}$	$p^\star_{310}$ $p^\star_{311}$ $p^\star_{312}$ $p^\star_{313}$
Asymptomatic Addicts	$p^\star_{020}$ $p^\star_{021}$ $p^\star_{022}$ $p^\star_{023}$	$p^\star_{120}$ $p^\star_{121}$ $p^\star_{122}$ $p^\star_{123}$	$p^\star_{220}$ $p^\star_{221}$ $p^\star_{222}$ $p^\star_{223}$	$p^\star_{320}$ $p^\star_{321}$ $p^\star_{322}$ $p^\star_{323}$
Pre-AIDS Addicts	$p^\star_{030}$ $p^\star_{031}$ $p^\star_{032}$ $p^\star_{033}$	$p^\star_{130}$ $p^\star_{131}$ $p^\star_{132}$ $p^\star_{133}$	$p^\star_{230}$ $p^\star_{231}$ $p^\star_{232}$ $p^\star_{233}$	$p^\star_{330}$ $p^\star_{331}$ $p^\star_{332}$ $p^\star_{333}$

infectivity, hence $p^\star_{231} = 0$. In a similar fashion we should have that the viral load of a needle after use must be greater than or equal to the minimum of the viral load of the current user and the state of the needle prior to use. For example a needle in state one infectivity used by an addict in stage three infectivity cannot enter state two infectivity, hence $p^\star_{132} = 0$. Therefore using common sense we can reduce the set of unknown $p^\star_{ijk}$ terms from 64 down to 32. This leaves us with the addict-needle interactions shown in Table 5.2.

We expect the probability that a needle is left in a given infectious state will depend on the infectious state of the needle prior to use and the infectivity of the current user. Therefore intuitively there are a number of inequalities between the various $p^\star_{ijk}$ terms which should always be satisfied. For example we expect that a needle in state three infectivity prior to use will have a higher probability of being left in state one

Table 5.2: Addict-Needle Interactions with Common Sense Adjustments

Addict Infectivity	Needle Infectivity (prior to use)			
	Uninfectious	Acute	Asymptomatic	Pre-AIDS
Uninfectious Addicts	$p^\star_{000} = 1$ $p^\star_{001} = 0$ $p^\star_{002} = 0$ $p^\star_{003} = 0$	$p^\star_{100}$ $p^\star_{101}$ $p^\star_{102}$ $p^\star_{103}$	$p^\star_{200}$ $p^\star_{201} = 0$ $p^\star_{202}$ $p^\star_{203} = 0$	$p^\star_{300}$ $p^\star_{301} = 0$ $p^\star_{302}$ $p^\star_{303}$
Acute Addicts	$p^\star_{010}$ $p^\star_{011}$ $p^\star_{012}$ $p^\star_{013}$	$p^\star_{110} = 0$ $p^\star_{111} = 1$ $p^\star_{112} = 0$ $p^\star_{113} = 0$	$p^\star_{210} = 0$ $p^\star_{211}$ $p^\star_{212}$ $p^\star_{213}$	$p^\star_{310} = 0$ $p^\star_{311}$ $p^\star_{312} = 0$ $p^\star_{313}$
Asymptomatic Addicts	$p^\star_{020}$ $p^\star_{021} = 0$ $p^\star_{022}$ $p^\star_{023} = 0$	$p^\star_{120} = 0$ $p^\star_{121}$ $p^\star_{122}$ $p^\star_{123}$	$p^\star_{220} = 0$ $p^\star_{221} = 0$ $p^\star_{222} = 1$ $p^\star_{223} = 0$	$p^\star_{320} = 0$ $p^\star_{321} = 0$ $p^\star_{322}$ $p^\star_{323}$
Pre-AIDS Addicts	$p^\star_{030}$ $p^\star_{031} = 0$ $p^\star_{032}$ $p^\star_{033}$	$p^\star_{130} = 0$ $p^\star_{131}$ $p^\star_{132} = 0$ $p^\star_{133}$	$p^\star_{230} = 0$ $p^\star_{231} = 0$ $p^\star_{232}$ $p^\star_{233}$	$p^\star_{330} = 0$ $p^\star_{331} = 0$ $p^\star_{332} = 0$ $p^\star_{333} = 1$

infectivity after use by an addict in stage one infectivity than if the needle were in state two infectivity prior to use by the same addict. Thus $p^\star_{311} \geq p^\star_{211}$, and similarly $p^\star_{211} \geq p^\star_{011}$. We can determine the inequalities which should exist between the $p^\star_{ijk}$ terms by considering the following four events:

1. A needle is left in state one infectivity.

2. A needle is left in state three or state one infectivity.

3. A needle is left in state two or state three or state one infectivity.

4. A needle is left in state two or state three infectivity.

Examining each of these four events gives us the following relationships between the $p^\star_{ijk}$ terms:

$$p^\star_{311} \geq p^\star_{211} \geq p^\star_{011}, \tag{5.1}$$

$$p^{\star}_{131} \geq p^{\star}_{121} \geq p^{\star}_{101}, \tag{5.2}$$

$$p^{\star}_{121} + p^{\star}_{123} \geq p^{\star}_{101} + p^{\star}_{103} \geq p^{\star}_{303}, \tag{5.3}$$

$$p^{\star}_{211} + p^{\star}_{213} \geq p^{\star}_{011} + p^{\star}_{013} \geq p^{\star}_{033}, \tag{5.4}$$

$$p^{\star}_{121} + p^{\star}_{123} \geq p^{\star}_{323}, \tag{5.5}$$

$$p^{\star}_{211} + p^{\star}_{213} \geq p^{\star}_{233}, \tag{5.6}$$

$$p^{\star}_{233} \geq p^{\star}_{033}, \tag{5.7}$$

$$p^{\star}_{323} \geq p^{\star}_{303}, \tag{5.8}$$

$$p^{\star}_{011} + p^{\star}_{012} + p^{\star}_{013} \geq p^{\star}_{033} + p^{\star}_{032} \geq p^{\star}_{022}, \tag{5.9}$$

$$p^{\star}_{101} + p^{\star}_{103} + p^{\star}_{102} \geq p^{\star}_{303} + p^{\star}_{302} \geq p^{\star}_{202}. \tag{5.10}$$

For example the inequality $p^{\star}_{121} + p^{\star}_{123} \geq p^{\star}_{101} + p^{\star}_{103}$ says that the probability that an Acutely Infectious needle used by an initially Asymptomatic addict is left in either the Acutely Infectious or Pre-AIDS state exceeds the corresponding probability for an uninfected addict. We expect this to be true as the blood of an Asymptomatic infected addict contains more virus than that of an uninfected one, so if both use an initially Acutely Infectious needle, the probability that the level of virus left in the needle is greater than or equal to any given level (in this case the level for a Pre-AIDS state infectivity needle) is greater for the former addict than the latter. Therefore when examining our model we should ensure that the above inequalities are always satisfied. This has the advantage of restricting the parameter space which as we discuss shortly assists with deriving analytical results.

We now outline a convenient method of including needle cleaning implicitly into our addict-needle interaction structure. Cleaning is a very important feature in modelling the spread of HIV among intravenous drug users. If needles are shared then cleaning a needle, either prior to use or after use, provides the most direct method of preventing new infections. Each $p^{\star}_{ijk}$ term denotes the probability that a needle initially in state i is left in state k immediately after use by an addict in stage j, it is natural to include cleaning into this probability. So far we have considered $p^{\star}_{ijk}$ as the probability of reaching state k by the process of "blood mixing" only. Denoting p_{ijk} as the case where cleaning is also allowed we have the relationship,

$$\begin{aligned} p_{ijk} &= P((i,j) \to k \mid \text{no cleaning}) P(\text{no cleaning}) + P((i,j) \to k \mid \text{cleaning}) P(\text{cleaning}) \\ &= p^{\star}_{ijk}(1-\phi) + p^{\star}_{0jk}\phi. \end{aligned}$$

Hence it is easy to incorporate cleaning implicitly into our addict-needle interaction

structure by replacing $p^{\star}_{ijk}$ with p_{ijk}. Note that if $\sum_{k=0}^{3} p^{\star}_{ijk} = 1$ we also have that $\sum_{k=0}^{3} p_{ijk} = 1$ and the previous inequalities still hold if they hold for the $p^{\star}_{ijk}$ terms.

From initially starting with 64 unknown addict-needle interaction probabilities ($p^{\star}_{ijk}$'s) we have reduced this number down to 32 and have introduced 10 sets of inequalities between the various probabilities. While this has served to simplify greatly the addict-needle interaction structure we still have a large number of unknown parameters. Before we discuss a model based on this most general set of interaction assumptions we first examine a slightly more restrictive model. We can simplify our general addict-needle interaction assumptions by making the highly plausible assumption that an uncontaminated needle used once by an addict in stage i infectivity becomes a state i infectious needle. We have made this assumption in the Optimistic and Pessimistic Models and almost all previous work makes the analogous assumption that an infected addict using an uninfected needle always leaves the needle infected (Kaplan, 1989, Peterson et al., 1990, Kaplan and O'Keefe, 1993, Seitz and Müller, 1994). This implies that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$. Using the inequalities (5.1) and (5.7) we additionally have that $p^{\star}_{211} = p^{\star}_{311} = p^{\star}_{233} = 1$. This reduces the number of unknown p_{ijk} terms (recall that these include the effect of cleaning) in our model to a much more manageable 16 and gives us the revised set of addict-needle interaction probabilities shown in Table 5.3.

We can now use the terms in Table 5.3 to construct a model which can incorporate a variety of addict-needle interaction assumptions. It is obviously the case that by assuming that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$ we have narrowed the scope of our model. However this model is significantly more flexible than either the Optimistic or Pessimistic Models. Moreover as we discuss later the Optimistic and Pessimistic Models are important special cases of this (restricted) general model.

5.3 Model Derivation

We now derive the differential equations which define the spread of HIV among an intravenous drug addict population where addicts progress through three stages of infectivity and addicts and needles interact according to Table 5.3. The addict equations in this model are the same as those in the Optimistic and Pessimistic Models, hence

Table 5.3: Revised Addict-Needle Interactions

Addict Infectivity	Needle Infectivity (prior to use)			
	Uninfectious	Acute	Asymptomatic	Pre-AIDS
Uninfectious Addicts	$p_{000} = 1$ $p_{001} = 0$ $p_{002} = 0$ $p_{003} = 0$	p_{100} p_{101} p_{102} p_{103}	p_{200} $p_{201} = 0$ p_{202} $p_{203} = 0$	p_{300} $p_{301} = 0$ p_{302} p_{303}
Acute Addicts	$p_{010} = 0$ $p_{011} = 1$ $p_{012} = 0$ $p_{013} = 0$	$p_{110} = 0$ $p_{111} = 1$ $p_{112} = 0$ $p_{113} = 0$	$p_{210} = 0$ $p_{211} = 1$ $p_{212} = 0$ $p_{213} = 0$	$p_{310} = 0$ $p_{311} = 1$ $p_{312} = 0$ $p_{313} = 0$
Asymptomatic Addicts	$p_{020} = 0$ $p_{021} = 0$ $p_{022} = 1$ $p_{023} = 0$	$p_{120} = 0$ p_{121} p_{122} p_{123}	$p_{220} = 0$ $p_{221} = 0$ $p_{222} = 1$ $p_{223} = 0$	$p_{320} = 0$ $p_{321} = 0$ p_{322} p_{323}
Pre-AIDS Addicts	$p_{030} = 0$ $p_{031} = 0$ $p_{032} = 0$ $p_{033} = 1$	$p_{130} = 0$ p_{131} $p_{132} = 0$ p_{133}	$p_{230} = 0$ $p_{231} = 0$ $p_{232} = 0$ $p_{233} = 1$	$p_{330} = 0$ $p_{331} = 0$ $p_{332} = 0$ $p_{333} = 1$

we have that

$$\frac{d\pi_1}{dt} = \Big(1 - \sum_{i=1}^{3} \pi_i\Big)\lambda(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1-\phi) - (\mu + \delta_1)\pi_1,$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu + \delta_2)\pi_2,$$

and $$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu + \delta_3)\pi_3.$$

Moving on to infectious needles we have that the number of state one infectious needles at time $t + \Delta t$

$=$ {number of state one infectious needles at time t}

+{number of addict-needle interactions in $[t, t + \Delta t)$

which result in a non-state one needle becoming state one}

$-\{$number of addict-needle interactions in $[t, t+\Delta t)$

which result in a state one needle becoming non-state one$\}$

$-\{$number of state one needles exchanged in $[t, t+\Delta t)\}$.

We can use Table 5.3 to identify which addict-needle interactions we require and then match these with the correct types of needles and addicts. For example we match p_{211} with the number of state two needles at time t, $m\beta_2(t)$ and the probability that a randomly selected needle is used by a stage one addict in $[t, t+\Delta t)$, $\lambda\gamma\Delta t\pi_1(t)$. Matching all the relevant p_{ijk} terms in this manner and writing $\pi = \pi_1 + \pi_2 + \pi_3$ gives us

$$\begin{aligned} m\beta_1(t+\Delta t) = m\beta_1(t) &+ m\lambda\Delta t\gamma\pi_1(t)(1-\beta(t)) \\ &+ m\lambda\Delta t\gamma\beta_2(t)\pi_1(t) \\ &+ m\lambda\Delta t\gamma\beta_3(t)\pi_1(t) \\ &- m\lambda\Delta t\gamma\beta_1(t)(1-\pi(t))(1-p_{101}) \\ &- m\lambda\Delta t\gamma\beta_1(t)\pi_2(t)(1-p_{121}) \\ &- m\lambda\Delta t\gamma\beta_1(t)\pi_3(t)(1-p_{131}) \\ &- m\beta_1(t)\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m\beta_1(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\begin{aligned} \frac{d\beta_1}{dt} &= \lambda\gamma\pi_1(1-\beta) + \lambda\gamma\beta_2\pi_1 + \lambda\gamma\beta_3\pi_1 - \lambda\gamma\beta_1(1-\pi)(1-p_{101}) \\ &\quad -\lambda\gamma\beta_1\pi_2(1-p_{121}) - \lambda\gamma\beta_1\pi_3(1-p_{131}) - \beta_1\tau. \end{aligned}$$

We have that the number of state two infectious needles at time $t+\Delta t$

$=$ $\{$number of state two infectious needles at time $t\}$

$+\{$number of addict-needle interactions in $[t, t+\Delta t)$

which result in a non-state two needle becoming state two$\}$

$-\{$number of addict-needle interactions in $[t, t+\Delta t)$

which result in a state two needle becoming non-state two$\}$

$-\{$number of state two needles exchanged in $[t, t+\Delta t)\}$.

As for state one needles we use Table 5.3 to identify which addict-needle interactions we require and then match these with the correct types of needles and addicts. This

gives us

$$\begin{aligned} m\beta_2(t+\Delta t) = m\beta_2(t) &+ m\lambda\Delta t\gamma\pi_2(t)(1-\beta(t)) \\ &+ m\lambda\Delta t\gamma\beta_1(t)\pi_2(t)p_{122} \\ &+ m\lambda\Delta t\gamma\beta_3(t)\pi_2(t)p_{322} \\ &+ m\lambda\Delta t\gamma\beta_1(t)(1-\pi(t))p_{102} \\ &+ m\lambda\Delta t\gamma\beta_3(t)(1-\pi(t))p_{302} \\ &- m\lambda\Delta t\gamma\beta_2(t)(1-\pi(t))(1-p_{202}) \\ &- m\lambda\Delta t\gamma\beta_2(t)\pi_1(t) \\ &- m\lambda\Delta t\gamma\beta_2(t)\pi_3(t) \\ &- m\beta_2(t)\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m\beta_2(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\begin{aligned} \frac{d\beta_2}{dt} &= \lambda\gamma\pi_2(1-\beta) + \lambda\gamma\beta_1\pi_2 p_{122} + \lambda\gamma\beta_3\pi_2 p_{322} + \lambda\gamma\beta_1(1-\pi)p_{102} \\ &\quad +\lambda\gamma\beta_3(1-\pi)p_{302} - \lambda\gamma\beta_2(1-\pi)(1-p_{202}) - \lambda\gamma\beta_2\pi_1 \\ &\quad -\lambda\gamma\beta_2\pi_3 - \beta_2\tau. \end{aligned}$$

We have that the number of state three infectious needles at time $t+\Delta t$

$$\begin{aligned} &= \{\text{number of state three infectious needles at time } t\} \\ &\quad +\{\text{number of addict-needle interactions in } [t, t+\Delta t) \\ &\quad \text{which result in a non-state three needle becoming state three}\} \\ &\quad -\{\text{number of addict-needle interactions in } [t, t+\Delta t) \\ &\quad \text{which result in a state three needle becoming non-state three}\} \\ &\quad -\{\text{number of state three needles exchanged in } [t, t+\Delta t)\}. \end{aligned}$$

Matching all the relevant p_{ijk} terms gives

$$\begin{aligned} m\beta_3(t+\Delta t) = m\beta_3(t) &+ m\lambda\Delta t\gamma\pi_3(t)(1-\beta(t)) \\ &+ m\lambda\Delta t\gamma\beta_1(t)\pi_2(t)p_{123} \\ &+ m\lambda\Delta t\gamma\beta_1(t)\pi_3(t)p_{133} \\ &+ m\lambda\Delta t\gamma\beta_2(t)\pi_3(t) \\ &+ m\lambda\Delta t\gamma\beta_1(t)(1-\pi(t))p_{103} \end{aligned}$$

$$
\begin{aligned}
&- m\lambda\Delta t\gamma\beta_3(t)(1-\pi(t))(1-p_{303}) \\
&- m\lambda\Delta t\gamma\beta_3(t)\pi_1(t) \\
&- m\lambda\Delta t\gamma\beta_3(t)\pi_2(t)(1-p_{323}) \\
&- m\beta_3(t)\tau\Delta t + o(\Delta t).
\end{aligned}
$$

Subtracting $m\beta_3(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$
\begin{aligned}
\frac{d\beta_3}{dt} &= \lambda\gamma\pi_3(1-\beta) + \lambda\gamma\beta_1\pi_2 p_{123} + \lambda\gamma\beta_1\pi_3 p_{133} + \lambda\gamma\beta_2\pi_3 \\
&\quad +\lambda\gamma\beta_1(1-\pi)p_{103} - \lambda\gamma\beta_3(1-\pi)(1-p_{303}) - \lambda\gamma\beta_3\pi_1 \\
&\quad -\lambda\gamma\beta_3\pi_2(1-p_{323}) - \beta_3\tau.
\end{aligned}
$$

Hence the system of differential equations which describes the spread of the disease is:

$$
\frac{d\pi_1}{dt} = \Big(1-\sum_{i=1}^{3}\pi_i\Big)\lambda(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)(1-\phi) - (\mu+\delta_1)\pi_1, \tag{5.11}
$$

$$
\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2, \tag{5.12}
$$

$$
\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3, \tag{5.13}
$$

$$
\begin{aligned}
\frac{d\beta_1}{dt} &= \lambda\gamma\pi_1(1-\beta) + \lambda\gamma\beta_2\pi_1 + \lambda\gamma\beta_3\pi_1 - \lambda\gamma\beta_1(1-\pi)(1-p_{101}) \\
&\quad -\lambda\gamma\beta_1\pi_2(1-p_{121}) - \lambda\gamma\beta_1\pi_3(1-p_{131}) - \beta_1\tau,
\end{aligned} \tag{5.14}
$$

$$
\begin{aligned}
\frac{d\beta_2}{dt} &= \lambda\gamma\pi_2(1-\beta) + \lambda\gamma\beta_1\pi_2 p_{122} + \lambda\gamma\beta_3\pi_2 p_{322} + \lambda\gamma\beta_1(1-\pi)p_{102} \\
&\quad +\lambda\gamma\beta_3(1-\pi)p_{302} - \lambda\gamma\beta_2(1-\pi)(1-p_{202}) - \lambda\gamma\beta_2\pi_1 \\
&\quad -\lambda\gamma\beta_2\pi_3 - \beta_2\tau,
\end{aligned} \tag{5.15}
$$

and

$$
\begin{aligned}
\frac{d\beta_3}{dt} &= \lambda\gamma\pi_3(1-\beta) + \lambda\gamma\beta_1\pi_2 p_{123} + \lambda\gamma\beta_1\pi_3 p_{133} + \lambda\gamma\beta_2\pi_3 \\
&\quad +\lambda\gamma\beta_1(1-\pi)p_{103} - \lambda\gamma\beta_3(1-\pi)(1-p_{303}) - \lambda\gamma\beta_3\pi_1 \\
&\quad -\lambda\gamma\beta_3\pi_2(1-p_{323}) - \beta_3\tau,
\end{aligned} \tag{5.16}
$$

with suitable initial conditions: $0 \le \pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$, $\beta_1(0)$, $\beta_2(0)$, $\beta_3(0)$, $\pi_1(0) + \pi_2(0) + \pi_3(0) \le 1$ and $\beta_1(0) + \beta_2(0) + \beta_3(0) \le 1$. We shall refer to the above model as the Restricted General Model.

5.3.1 Complete Generality Model

We believe that the model in eqns (5.11)-(5.16) is interesting and useful in its own right and can be used to examine the impact of a three stage infectious period and differing addict-needle interaction assumptions. However we have made the simplifying assumption that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$. As mentioned above almost all previous work makes an analogous assumption, however this may be more pessimistic than would reasonably be expected since an uncontaminated needle may contain an amount of uninfectious blood rather than being completely devoid of any fluid. This uninfectious blood could serve to dilute the blood of the next infectious user and therefore $p^{\star}_{0jj} < 1$ for $j = 1, 2, 3$. In addition it may be that addicts rinse unused needles with water prior to use which would have a similar dilution effect. Hence it is more general and probably more realistic to simply assume that $p^{\star}_{0jj} < 1$ for $j = 1, 2, 3$. An additional reason for relaxing the assumption $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$ is that this causes an inconsistency in our model. We have used a general "blood-mixing" structure for interactions between infectious needles and infectious addicts but taken the extreme approach of full flushing in interactions between infectious addicts and uninfectious needles. In other words we have not allowed for the dilution effect in the latter interactions. We now state an extension of the previous model which assumes a general "blood-mixing" structure in all addict-needle interactions.

If we relax the assumption that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$ then using the inequalities in eqns (5.1) and (5.7) we no longer have that $p^{\star}_{211} = p^{\star}_{311} = p^{\star}_{233} = 1$. We can still incorporate cleaning implicitly into the p_{ijk} terms however care is required as it is now the case that a number of p_{ijk} terms (after cleaning is included) which were previously equal to zero are no longer zero. For example p_{120} (with cleaning) is not necessarily zero since if a state one needle is cleaned prior to use by a stage two addict then this needle could become uncontaminated since $p^{\star}_{020}$ (excluding cleaning) may now be strictly positive. Similarly terms like p_{111} may not equal one. It is straightforward to derive the equations which define this more general model in a similar fashion to the model in eqns (5.11)-(5.16). The equations which define the fully general model are:

$$\frac{d\pi_1}{dt} = \Big(1 - \sum_{i=1}^{3} \pi_i\Big)\lambda(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1-\phi) - \pi_1(\mu + \delta_1), \tag{5.17}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu + \delta_2)\pi_2, \tag{5.18}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3, \tag{5.19}$$

$$\frac{d\beta_1}{dt} = \lambda\gamma\pi_1(1-\beta)p_{011} + \lambda\gamma\beta_2\pi_1p_{211} + \lambda\gamma\beta_3\pi_1p_{311} - \lambda\gamma\beta_1(1-\pi)(1-p_{101})$$

$$-\lambda\gamma\beta_1\pi_2(1-p_{121}) - \lambda\gamma\beta_1\pi_3(1-p_{131}) - \lambda\gamma\beta_1\pi_1(1-p_{111}) - \beta_1\tau, \tag{5.20}$$

$$\frac{d\beta_2}{dt} = \lambda\gamma\pi_2(1-\beta)p_{022} + \lambda\gamma\beta_1\pi_2p_{122} + \lambda\gamma\beta_3\pi_2p_{322} + \lambda\gamma\beta_1(1-\pi)p_{102}$$

$$+\lambda\gamma\beta_3(1-\pi)p_{302} + \lambda\gamma\pi_1(1-\beta)p_{012} + \lambda\gamma\pi_3(1-\beta)p_{032} + \lambda\gamma\beta_1\pi_1p_{112}$$

$$+\lambda\gamma\beta_1\pi_3p_{132} + \lambda\gamma\beta_3\pi_1p_{312} + \lambda\gamma\beta_3\pi_3p_{332} - \lambda\gamma\beta_2(1-\pi)(1-p_{202})$$

$$-\lambda\gamma\beta_2\pi_1(1-p_{212}) - \lambda\gamma\beta_2\pi_2(1-p_{222}) - \lambda\gamma\beta_2\pi_3(1-p_{232}) - \beta_2\tau, \tag{5.21}$$

and

$$\frac{d\beta_3}{dt} = \lambda\gamma\pi_3(1-\beta)p_{033} + \lambda\gamma\beta_1\pi_2p_{123} + \lambda\gamma\beta_1\pi_3p_{133} + \lambda\gamma\beta_2\pi_1p_{213} + \lambda\gamma\beta_2\pi_3p_{233}$$

$$+\lambda\gamma\beta_1(1-\pi)p_{103} + \lambda\gamma\pi_1(1-\beta)p_{013} + \lambda\gamma\beta_1\pi_1p_{113} - \lambda\gamma\beta_3(1-\pi)(1-p_{303})$$

$$-\lambda\gamma\beta_3\pi_1(1-p_{313}) - \lambda\gamma\beta_3\pi_2(1-p_{323}) - \lambda\gamma\beta_3\pi_3(1-p_{333}) - \beta_3\tau, \tag{5.22}$$

again with suitable initial conditions: $0 \leq \pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$, $\beta_1(0)$, $\beta_2(0)$, $\beta_3(0)$, $\pi_1(0)+\pi_2(0)+\pi_3(0) \leq 1$ and $\beta_1(0)+\beta_2(0)+\beta_3(0) \leq 1$. The system of eqns (5.17)-(5.22) could be considered a generalisation of a single stage infectivity model similar to the Kaplan and O'Keefe Model but where an uninfectious needle is not necessarily left infectious by an infectious addict.

5.4 The Optimistic, Pessimistic and General Models

Having derived two general mixing models, the General Model and the Restricted General Model, we now briefly discuss the relationship between these models and the Optimistic and Pessimistic Models. It is obvious that the Optimistic and Pessimistic Models are special cases of these general models since we can choose $p^{\star}_{ijk}$ terms to give the same addict-needle interactions assumptions as in these models. When we constructed the Restricted General Model we made the simplifying assumption that an uninfectious needle used once by an addict in stage i infectivity becomes a state i infectious needle. This assumption implies that the Optimistic and Pessimistic Models are

in fact the best and worst cases of this general mixing model. In other words we intuitively expect that for the same choice of non-$p^\star_{ijk}$ parameters, the long term prevalence of disease in the Restricted General Model will lie in between that of the Optimistic Model and Pessimistic Model for *any* choice of $p^\star_{ijk}$'s such that the inequalities in eqns (5.1)-(5.10) are satisfied. The easiest way to see this is to examine Table 5.4 in Section 5.7. In the $p^\star_{ijk}$'s which we are free to adjust (the number of which is constrained due to the assumption that $p^\star_{011} = p^\star_{022} = p^\star_{033} = 1$) we find that the most infectious outcome always occurs with probability one in the Pessimistic Model and the least infectious outcome always occurs with probability one in the Optimistic Model.

This convenient relationship between the Restricted General Model and the Optimistic and Pessimistic Models suggests that these simpler models have practical use as lower and upper bounds for the spread of disease under the assumption of three stage infectivity. Note however that without the specific assumption that $p^\star_{011} = p^\star_{022} = p^\star_{033} = 1$ this connection between the models does not hold. Intuitively the Pessimistic Model still represents an upper bound for the spread of disease but the Optimistic Model no longer represents a lower bound. To see this consider for example the interaction between a stage one addict and an uninfectious needle. In the Optimistic Model (and the Restricted General Model) this needle will be left in state one infectivity with probability one. However in the General Model where we have dropped the condition $p^\star_{011} = p^\star_{022} = p^\star_{033} = 1$ it is now possible for this uninfectious needle to be left in the lower infectious states of state two or state three or even remain uninfectious. Hence we no longer expect the Optimistic Model to represent a lower bound.

5.5 The Basic Reproductive Number

As in the previous models we are interested in deriving the basic reproductive number for the General Model. The computation of R_0 in this model is more complicated than for the earlier models. In the Optimistic and Pessimistic Models an infectious needle could only remain in the same infectious state or be rendered virus free by an uncontaminated addict. In the General Model uncontaminated addicts can lower the viral load of a contaminated needle so that it enters other less infectious states before being rendered virus free.

As usual, consider a single newly infected addict entering a population where everyone else is susceptible and all needles are uninfected. We wish to calculate the expected

number of secondary cases produced by this single infected individual. Addicts inject at rate λ per unit time and on average spend $1/(\mu+\delta_1)$ time units in stage one infectivity. Hence

$$\frac{\lambda p_{011}}{\mu+\delta_1}, \quad \frac{\lambda p_{012}}{\mu+\delta_1} \quad \text{and} \quad \frac{\lambda p_{013}}{\mu+\delta_1}$$

are the respective expected numbers of needles an addict leaves in state one, two and three infectivity during his or her stage one lifetime. An addict progresses from stage one to stage two with probability $\delta_1/(\mu+\delta_1)$ and spends on average $1/(\mu+\delta_2)$ time units in this stage. On average an addict leaves

$$\frac{\lambda\delta_1 p_{022}}{(\mu+\delta_1)(\mu+\delta_2)},$$

needles in state two infectivity during his or her stage two lifetime. An addict progresses from stage two to stage three with probability $\delta_2/(\mu+\delta_2)$ and spends on average $1/(\mu+\delta_3)$ time units in this stage. Hence

$$\frac{\lambda\delta_1\delta_2 p_{032}}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)} \quad \text{and} \quad \frac{\lambda\delta_1\delta_2 p_{033}}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}$$

are the respective expected numbers of needles an addict leaves in state two and three infectivity during his or her stage three lifetime. We now determine how many infections are caused by each type of infectious needle until it is rendered virus free, where a needle can pass through states of lower infectivity on its way to becoming virus free. We have that

$$E(Y) \;=\; E(Y\,|\,X_1)P(X_1) + E(Y\,|\,X_2)P(X_2),$$

where Y is the number of addicts infected by a single needle, X_1 is the event that the needle is exchanged before the next injection and X_2 is the event that the needle is still infectious at the next injection. If the needle is exchanged prior to the next injection then the infected needle has infected zero addicts, thus $E(Y\,|\,X_1) = 0$. The event X_2 corresponds to the needle being used rather than exchanged prior to use. The probability of this event is $\lambda\gamma/(\lambda\gamma+\tau)$, hence

$$E(Y) = E(Y\,|\,X_2)\frac{\lambda\gamma}{\lambda\gamma+\tau}. \tag{5.23}$$

We explore $E(Y\,|\,X_2)$ by conditioning on the next event, that of a susceptible addict injecting with an infectious needle, we now first assume that this needle is in state two infectivity. This event has four outcomes given by pairs of combinations of the following two events. The addict is infected by the needle with probability $\alpha_2(1-\phi)$ or remains

susceptible with probability $1-\alpha_2(1-\phi)$, and also the addict can leave the needle in state two infectivity with probability p_{202} or leave the needle virus free with probability p_{200}. As discussed by Greenhalgh and Hay (1997) independence of the events that an addict is infected by the needle and the needle is flushed is not necessary. Let E_2 denote the unconditional total expected number of addicts infected from a needle initially in state two infectivity. Using eqn (5.23) we have that

$$\begin{aligned} E_2 &= \frac{\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected and needle left in state 2)}(1+E_2) \\ &\quad + \text{P(sus. addict infected and needle left virus free)} \\ &\quad + \text{P(sus. addict not infected and needle left in state 2)}E_2\Big], \\ &= \frac{\lambda\gamma}{\lambda\gamma+\tau}\Big[\text{P(sus. addict infected)} + \text{P(needle left in state 2)}E_2\Big], \\ &= \frac{1}{1+\hat{\tau}}[(1-\phi)\alpha_2 + p_{202}E_2], \end{aligned}$$

where $\hat{\tau} = \tau/\lambda\gamma$. Hence solving for E_2 gives us

$$E_2 = \frac{(1-\phi)\alpha_2}{1+\hat{\tau}-p_{202}}.$$

Hence E_2 is the probability that an addict is infected at the next event $((1-\phi)\alpha_2)/(1+\hat{\tau})$, divided by the probability that a state two needle is rendered virus free at the next event, $1-(p_{202}/(1+\hat{\tau}))$. Note that whilst we have included cleaning implicitly into the p_{ijk} terms we still feature ϕ explicitly. Cleaning has two effects, firstly if susceptible addicts clean needles prior to use then they increase the chance that a previously infectious needle is rendered virus free. We have implicitly incorporated this effect in our definition of p_{ijk}. The second effect of cleaning is to reduce the probability of HIV transmission from the infectious needle to the susceptible addict, thus cleaning reduces the probability of HIV transmission from α_2 to $\alpha_2(1-\phi)$, and correspondingly raises the probability that HIV is not transmitted from $1-\alpha_2$ to $1-\alpha_2(1-\phi)$.

Moving on to state three needles, we again wish to compute the expected number of addicts infected by a single infectious needle where now the needle is initially in state three infectivity. This time our model allows a susceptible addict to leave a needle initially in state three infectivity in state three or state two infectivity, as well as rendering it virus-free. Let E_3 denote the unconditional total expected number of

addicts infected given the needle is initially in state three infectivity. Then arguing similarly to before we have

$$\begin{aligned} E_3 &= \frac{\lambda\gamma}{\lambda\gamma+\tau}\Big[\mathrm{P}(\text{sus. addict infected}) + \mathrm{P}(\text{needle left in state two})E_2 \\ &\qquad + \mathrm{P}(\text{needle left in state three})E_3\Big], \\ &= \frac{1}{1+\hat{\tau}}[\alpha_3(1-\phi) + p_{302}E_2 + p_{303}E_3]. \end{aligned}$$

Hence

$$\begin{aligned} E_3 &= \frac{\alpha_3(1-\phi)}{1+\hat{\tau}-p_{303}} + \frac{p_{302}E_2}{1+\hat{\tau}-p_{303}}, \\ &= \frac{\alpha_3(1-\phi)}{1+\hat{\tau}-p_{303}} + \frac{p_{302}}{1+\hat{\tau}-p_{303}}\frac{\alpha_2(1-\phi)}{1+\hat{\tau}-p_{202}}, \end{aligned}$$

using our previously derived expression for E_2.

Moving on to needles in state one infectivity, let E_1 denote the corresponding unconditional total expected number of addicts infected given that the needle is initially in state one infectivity. Then arguing as above

$$E_1 = \frac{1}{1+\hat{\tau}}\Big[\alpha_1(1-\phi) + p_{102}E_2 + p_{103}E_3 + p_{101}E_1\Big].$$

Solving for E_1 gives us

$$E_1 = \frac{\alpha_1(1-\phi)}{1+\hat{\tau}-p_{101}} + \frac{p_{102}E_2}{1+\hat{\tau}-p_{101}} + \frac{p_{103}E_3}{1+\hat{\tau}-p_{101}}.$$

Substituting our previously derived expressions for E_2 and E_3 we deduce that

$$\begin{aligned} E_1 &= \frac{(1-\phi)\alpha_1}{1+\hat{\tau}-p_{101}} + \frac{p_{103}}{1-p_{101}+\hat{\tau}}\left[\frac{(1-\phi)\alpha_3}{1+\hat{\tau}-p_{303}} + \frac{p_{302}}{(1-p_{303}+\hat{\tau})}\frac{(1-\phi)\alpha_2}{(1+\hat{\tau}-p_{202})}\right] \\ &\quad + \frac{p_{102}}{(1-p_{101}+\hat{\tau})}\frac{(1-\phi)\alpha_2}{(1+\hat{\tau}-p_{202})}. \end{aligned}$$

At the start of this section we computed the expected number of state one, state two and state three infectious needles an addict creates during his or her entire infectious lifetime, we now have additionally the expected number of addicts infected by each of these three types of infectious needles. Putting these expectations together gives $R_0 =$

$$\begin{aligned} &\frac{\lambda(1-\phi)p_{011}}{(\mu+\delta_1)}\left\{\frac{\alpha_1}{1+\hat{\tau}-p_{101}} + \frac{p_{103}}{1-p_{101}+\hat{\tau}}\left[\frac{\alpha_3}{1+\hat{\tau}-p_{303}} + \frac{p_{302}}{(1-p_{303}+\hat{\tau})}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}\right]\right. \\ &\left. + \quad \frac{p_{102}}{(1-p_{101}+\hat{\tau})}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}\right\} \end{aligned}$$

$$+ \quad \frac{\lambda(1-\phi)}{(\mu+\delta_1)}\left[p_{012}+\frac{\delta_1 p_{022}}{(\mu+\delta_2)}+\frac{\delta_1\delta_2 p_{032}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}$$

$$+ \quad \frac{\lambda(1-\phi)}{(\mu+\delta_1)}\left[p_{013}+\frac{\delta_1\delta_2 p_{033}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\left[\frac{\alpha_3}{1+\hat{\tau}-p_{303}}+\frac{p_{302}}{(1-p_{303}+\hat{\tau})}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}\right].$$

For the Restricted General Model this expression simplifies down to

$$\frac{\lambda(1-\phi)}{(\mu+\delta_1)}\left\{\frac{\alpha_1}{1+\hat{\tau}-p_{101}}+\frac{p_{103}}{1-p_{101}+\hat{\tau}}\left[\frac{\alpha_3}{1+\hat{\tau}-p_{303}}+\frac{p_{302}}{(1-p_{303}+\hat{\tau})}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}\right]\right.$$

$$\left. + \quad \frac{p_{102}}{(1-p_{101}+\hat{\tau})}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}\right\}+\frac{\lambda(1-\phi)}{(\mu+\delta_1)}\frac{\delta_1}{(\mu+\delta_2)}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}$$

$$+ \quad \frac{\lambda(1-\phi)}{(\mu+\delta_1)}\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\left[\frac{\alpha_3}{1+\hat{\tau}-p_{303}}+\frac{p_{302}}{(1-p_{303}+\hat{\tau})}\frac{\alpha_2}{(1+\hat{\tau}-p_{202})}\right].$$

Not surprisingly the expressions for R_0 in the General Model and the Restricted General Model are more complicated than those for the Optimistic and Pessimistic Models. However it is easy to see that these expressions collapse down to the expressions of R_0 for the Optimistic and Pessimistic Models given the appropriate choice of p_{ijk} terms. We now move on to studying the analytical properties of our model and demonstrate the importance of the above expression for R_0.

5.6 Analytical Results

We have derived two models, that defined by eqns (5.11)-(5.16) and its extension to complete generality defined by eqns (5.17)-(5.22). We now show a number of analytical properties of these systems of equations. We assume that all model parameters (apart from ϕ and the p_{ijk} parameters) are strictly positive, ϕ is strictly less than one and $\alpha_1 > \alpha_3 > \alpha_2$. In addition we assume that $p_{0jj} > 0$ for $j = 1,2,3$ to avoid technical complications where disease may die out in one or more classes of needles but the disease is still endemic. This condition says that an addict can leave an uninfectious needle in their own infectious state so seems entirely reasonable. In some of the following results we also require that $p_{011} \geq \max(p_{211}, p_{311})$, $p_{011}+p_{013} \geq p_{211}+p_{213}$ and $p_{033} \geq p_{233}$, these are required for technical reasons rather than biological reality. These inequalities are always true for the simpler model in eqns (5.11)-(5.16) and together with inequalities (5.1)-(5.10) implies that $p_{011} = p_{211} = p_{311}$, $p_{013} = p_{213}$ and $p_{033} = p_{233}$.

Theorem 5.1 *If $R_0 \leq 1$ the system of equations (5.17)-(5.22) has a unique equilibrium solution where the disease has died out in both addicts and needles. If in addition,*

$p_{011} \geq \max(p_{211}, p_{311})$, $p_{011} + p_{013} \geq p_{211} + p_{213}$ and $p_{033} \geq p_{233}$ then whatever the initial state the disease will die out in all addicts and all needles. If $R_0 > 1$ there is still the equilibrium where the disease has died out and this is unstable, in addition there now also exists at least one endemic equilibrium solution with disease present in each class of addict and each class of needle.

Proof.

The proof of this theorem requires a number of stages and supporting arguments. We first define a function $\mathcal{L}(\mathbf{x})$ which is of central importance in our proof. For simplicity of notation we write $\pi_{1+3} = \pi_1 + \pi_3$, $\pi = \pi_1 + \pi_2 + \pi_3$, $\beta_{1+3} = \beta_1 + \beta_3$ and $\beta = \beta_1 + \beta_2 + \beta_3$. Let $\mathbf{x} = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_{1+3}, \beta)$, $\mathcal{L}(\mathbf{x})$ is a scalar function of the six components in $\mathbf{x}$ given by $\mathcal{L}(\mathbf{x})$

$$= \frac{\zeta_1}{\mu+\delta_1}\beta_1 + \frac{\zeta_2}{\mu+\delta_1}\beta_{1+3} + \frac{\zeta_3}{\mu+\delta_1}\beta + \frac{\gamma}{(\mu+\delta_1)}\left[\pi_1 + \frac{\lambda\pi_3\left(p_{033}(\zeta_2+\zeta_3) + p_{032}\zeta_3\right)}{(\mu+\delta_3)}\right.$$

$$\left. + \frac{\lambda\pi_2}{(\mu+\delta_2)}\left(p_{022}\zeta_3 + \left(p_{033}(\zeta_2+\zeta_3) + p_{032}\zeta_3\right)\frac{\delta_2}{\mu+\delta_3}\right)\right]. \tag{5.24}$$

Here

$$\zeta_1 = \frac{\alpha_1(1-\phi)}{1-p_{101}+\hat{\tau}} - \frac{1}{(1-p_{101}+\hat{\tau})}\left\{\frac{\alpha_2(1-\phi)}{1-p_{202}+\hat{\tau}}(1-p_{101}+\hat{\tau}-p_{102}-p_{103})\right.$$

$$\left. + \left[\alpha_3(1-\phi) - \frac{\alpha_2(1-\phi)(1-p_{303}+\hat{\tau}-p_{302})}{1-p_{202}+\hat{\tau}}\right]\left[\frac{1-p_{101}+\hat{\tau}-p_{103}}{1-p_{303}+\hat{\tau}}\right]\right\}, \tag{5.25}$$

$$\zeta_2 = \frac{\alpha_3(1-\phi)}{1-p_{303}+\hat{\tau}} - \frac{\alpha_2(1-\phi)(1-p_{303}+\hat{\tau}-p_{302})}{(1-p_{303}+\hat{\tau})(1-p_{202}+\hat{\tau})}, \tag{5.26}$$

$$\text{and } \zeta_3 = \frac{\alpha_2(1-\phi)}{1-p_{202}+\hat{\tau}}. \tag{5.27}$$

Lemma 5.1 *If $p_{011} \geq \max(p_{211}, p_{311})$, $p_{011}+p_{013} \geq p_{211}+p_{213}$ and $p_{033} \geq p_{233}$ then the function $\mathcal{L}(\mathbf{x})$ is a weak Lyapunov function for the system defined by eqns (5.17)-(5.22).*

Proof.

Using Theorem 10.1 in Jordan and Smith (1987), we require to show that $\mathcal{L}(\mathbf{x})$ and its partial derivatives are continuous, $\mathcal{L}(\mathbf{x})$ is positive definite and $d\mathcal{L}(\mathbf{x})/dt$ is negative semidefinite. It is obvious that $\mathcal{L}(\mathbf{x})$ and its partial derivatives are continuous. We now show that $\mathcal{L}(\mathbf{x})$ is positive definite. It is obvious that $\zeta_3 > 0$ always. Since $\alpha_3 > \alpha_2$

and $p_{303} + p_{302} \geq p_{202}$ (from inequality (5.10)) then we have that $\zeta_2 > 0$ always. We also have that

$$\begin{aligned}
\zeta_1 &\geq \frac{\alpha_1(1-\phi)}{1-p_{101}+\hat{\tau}} - \frac{1}{1-p_{101}+\hat{\tau}}\left\{\frac{\alpha_2(1-\phi)(1-p_{101}+\hat{\tau}-p_{102}-p_{103})}{1-p_{202}+\hat{\tau}} + \alpha_3(1-\phi)\right. \\
&\quad \left. -\frac{\alpha_2(1-\phi)(1-p_{303}-p_{302}+\hat{\tau})}{1-p_{202}+\hat{\tau}}\right\}, \qquad \text{since } p_{103}+p_{101} \geq p_{303} \text{ from eqn (5.3)}, \\
&\geq \frac{\alpha_1(1-\phi)}{1-p_{101}+\hat{\tau}} - \frac{\alpha_3(1-\phi)}{1-p_{101}+\hat{\tau}}, \qquad \text{since } p_{300} \geq p_{100} \text{ from eqn (5.10)}, \\
&> 0, \qquad \text{since } \alpha_1 > \alpha_3.
\end{aligned}$$

Hence the function $\mathcal{L}(\mathbf{x})$ is positive definite. Moving on to $d\mathcal{L}(\mathbf{x})/dt$ we have that

$$\frac{d\mathcal{L}}{dt} = \frac{\partial\mathcal{L}}{\partial\beta_1}\frac{d\beta_1}{dt} + \frac{\partial\mathcal{L}}{\partial\beta_{1+3}}\frac{\beta_{1+3}}{dt} + \frac{\partial\mathcal{L}}{\partial\beta}\frac{d\beta}{dt} + \frac{\partial\mathcal{L}}{\partial\pi_1}\frac{d\pi_1}{dt} + \frac{\partial\mathcal{L}}{\partial\pi_2}\frac{d\pi_2}{dt} + \frac{\partial\mathcal{L}}{\partial\pi_3}\frac{d\pi_3}{dt}. \quad (5.28)$$

Using eqns (5.17)-(5.22) it is straightforward to compute $d\mathcal{L}/dt$. After some simplification we have:

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\beta_1}\frac{d\beta_1}{dt} &= \frac{\zeta_1}{\mu+\delta_1}\left\{A - \pi_1\left[\beta_1(p_{011}+1-p_{111}) + \beta_2(p_{011}-p_{211})\right.\right. \\
&\quad \left.\left. +\beta_3(p_{011}-p_{311})\right]\lambda\gamma - \pi_2\beta_1(1-p_{121})\lambda\gamma - \pi_3\beta_1(1-p_{131})\lambda\gamma \right. \\
&\quad \left. -\beta_1\tau\pi\right\}, \qquad (5.29)
\end{aligned}$$

$$\text{where} \quad A = \lambda\gamma\pi_1 p_{011} - \lambda\gamma\beta_1(1-\pi)(1-p_{101}+\hat{\tau}), \qquad (5.30)$$

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\beta_{1+3}}\frac{d\beta_{1+3}}{dt} &= \frac{\zeta_2}{\mu+\delta_1}\left\{B - \lambda\gamma\pi_1\left[\beta_1(p_{011}+p_{013}+1-p_{111}-p_{113})\right.\right. \\
&\quad \left.\left. +\beta_2(p_{011}+p_{013}-p_{211}-p_{213}) + \beta_3(p_{011}+p_{013}+1-p_{313}-p_{311})\right]\right. \\
&\quad \left. -\lambda\gamma\pi_2\left[\beta_1(1-p_{121}-p_{123}) + \beta_3(1-p_{323})\right]\right. \\
&\quad \left. -\lambda\gamma\pi_3\left[\beta_1(p_{033}+1-p_{131}-p_{133}) + \beta_2(p_{033}-p_{233})\right.\right. \\
&\quad \left.\left. +\beta_3(p_{033}+1-p_{333})\right] - \beta_{1+3}\tau\pi\right\}, \qquad (5.31)
\end{aligned}$$

$$\begin{aligned}
\text{where} \quad B &= \lambda\gamma\left[\pi_1(p_{011}+p_{013}) + \pi_3 p_{033}\right] - (1-\pi)\beta_1(1-p_{101}-p_{103}+\hat{\tau})\lambda\gamma \\
&\quad -(1-\pi)\beta_3(1-p_{303}+\hat{\tau})\lambda\gamma, \qquad (5.32)
\end{aligned}$$

$$\frac{\partial\mathcal{L}}{\partial\beta}\frac{d\beta}{dt} = \frac{\zeta_3}{\mu+\delta_1}\left\{C - \lambda\gamma\pi_1\left[\beta_1(p_{011}+p_{012}+p_{013}+p_{110})\right.\right.$$

$$+\beta_2(p_{011}+p_{012}+p_{013}+p_{210})+\beta_3(p_{011}+p_{012}+p_{013}+p_{310})\Big]$$

$$-\lambda\gamma\pi_2\Big[\beta_1(p_{022}+p_{120})+\beta_2(p_{022}+p_{220})+\beta_3(p_{022}+p_{320})\Big]$$

$$-\lambda\gamma\pi_3\Big[\beta_1(p_{033}+p_{032}+p_{130})+\beta_2(p_{032}+p_{033}+p_{230})$$

$$+\beta_3(p_{032}+p_{033}+p_{330})\Big]-\beta\tau\pi\Big\}, \tag{5.33}$$

where $$C = \lambda\gamma\Big[\pi_1(p_{011}+p_{012}+p_{013})+\pi_2 p_{022}+\pi_3(p_{032}+p_{033})\Big]$$

$$-(1-\pi)\Big[\beta_1(p_{100}+\hat{\tau})+\beta_2(p_{200}+\hat{\tau})+\beta_3(p_{300}+\hat{\tau})\Big]\lambda\gamma, \tag{5.34}$$

$$\frac{\partial\mathcal{L}}{\partial\pi_1}\frac{d\pi_1}{dt} = \frac{\gamma}{\mu+\delta_1}\Big[\lambda(1-\phi)(1-\pi)(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)-(\mu+\delta_1)\pi_1\Big], \tag{5.35}$$

$$\frac{\partial\mathcal{L}}{\partial\pi_2}\frac{d\pi_2}{dt} = \frac{\lambda\gamma}{(\mu+\delta_1)(\mu+\delta_2)}\left(p_{022}\zeta_3+\Big(p_{033}(\zeta_2+\zeta_3)+p_{032}\zeta_3\Big)\frac{\delta_2}{(\mu+\delta_3)}\right)$$

$$\times[\delta_1\pi_1-(\mu+\delta_2)\pi_2], \tag{5.36}$$

and $$\frac{\partial\mathcal{L}}{\partial\pi_3}\frac{d\pi_3}{dt} = \frac{\lambda\gamma\Big(p_{033}(\zeta_2+\zeta_3)+p_{032}\zeta_3\Big)}{(\mu+\delta_1)(\mu+\delta_3)}[\delta_2\pi_2-(\mu+\delta_3)\pi_3]. \tag{5.37}$$

We can express $d\mathcal{L}/dt$ as

$$\frac{d\mathcal{L}}{dt} = \Big[\pi_1 p_{011}-\beta_1(1-\pi)(1-p_{101}+\hat{\tau})\Big]\frac{\lambda\gamma\zeta_1}{\mu+\delta_1}+\Big[\pi_1(p_{011}+p_{013})+\pi_3 p_{033}$$

$$-(1-\pi)(\beta_1(1-p_{101}+\hat{\tau}-p_{103})+\beta_3(1-p_{303}+\hat{\tau}))\Big]\frac{\lambda\gamma\zeta_2}{\mu+\delta_1}$$

$$+\Big[\pi_1(p_{011}+p_{012}+p_{013})+\pi_2 p_{022}+\pi_3(p_{032}+p_{033})$$

$$-(1-\pi)(\beta_1(1-p_{101}+\hat{\tau}-p_{102}-p_{103})+\beta_2(1-p_{202}+\hat{\tau})$$

$$+\beta_3(1-p_{303}+\hat{\tau}-p_{302}))\Big]\frac{\lambda\gamma\zeta_3}{\mu+\delta_1}+\frac{\gamma}{\mu+\delta_1}\Big\{\lambda(1-\phi)(1-\pi)(\alpha_1\beta_1+\alpha_2\beta_2$$

$$+\alpha_3\beta_3)-(\mu+\delta_1)\pi_1+\frac{\lambda\Big(p_{033}(\zeta_2+\zeta_3)+p_{032}\zeta_3\Big)}{(\mu+\delta_3)}[\delta_2\pi_2-(\mu+\delta_3)\pi_3]$$

$$+\frac{\lambda}{\mu+\delta_2}\left(p_{022}\zeta_3+\Big(p_{033}(\zeta_2+\zeta_3)+p_{032}\zeta_3\Big)\frac{\delta_2}{\mu+\delta_3}\right)[\delta_1\pi_1-(\mu+\delta_2)\pi_2]\Big\}$$

$$+A^-+B^-+C^-, \tag{5.38}$$

where $\frac{\partial \mathcal{L}}{\partial \beta_1}\frac{d\beta_1}{dt} = A + A^-$, $\frac{\partial \mathcal{L}}{\partial \beta_{1+3}}\frac{d\beta_{1+3}}{dt} = B + B^-$ and $\frac{\partial \mathcal{L}}{\partial \beta}\frac{d\beta}{dt} = C + C^-$, and note that A^-, B^- and C^- are all negative.

By construction we have that

$$\beta_1(1 - p_{101} + \hat{\tau})\zeta_1 + \Big[\beta_1(1 - p_{101} - p_{103} + \hat{\tau}) + \beta_3(1 - p_{303} + \hat{\tau})\Big]\zeta_2$$

$$+\Big[\beta_1(1 - p_{101} - p_{102} - p_{103} + \hat{\tau}) + \beta_2(1 - p_{202} + \hat{\tau}) + \beta_3(1 - p_{303} - p_{302} + \hat{\tau})\Big]\zeta_3,$$

$$= (1 - \phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3). \tag{5.39}$$

Hence eqn (5.38) simplifies down to

$$\begin{aligned}\frac{d\mathcal{L}}{dt} &= \gamma\pi_1\Bigg[\frac{\lambda}{\mu + \delta_1}\Big(p_{011}(\zeta_1 + \zeta_2 + \zeta_3) + p_{013}(\zeta_2 + \zeta_3) + p_{012}\zeta_3\Big)\\ &\quad + \frac{\lambda\delta_1}{(\mu + \delta_1)(\mu + \delta_2)}p_{022}\zeta_3 + \frac{\lambda\delta_1\delta_2}{(\mu + \delta_1)(\mu + \delta_2)(\mu + \delta_3)}\Big(p_{033}(\zeta_2 + \zeta_3) + p_{032}\zeta_3\Big)\\ &\quad - 1\Bigg] + A^- + B^- + C^-. \end{aligned}\tag{5.40}$$

After some simplification we have that

$$\zeta_2 + \zeta_3 = \frac{\alpha_3(1 - \phi)(1 - p_{202} + \hat{\tau}) + \alpha_2(1 - \phi)p_{302}}{(1 - p_{303} + \hat{\tau})(1 - p_{202} + \hat{\tau})}$$

and

$$\begin{aligned}\zeta_1 + \zeta_2 + \zeta_3 &= \frac{\alpha_1(1 - \phi)(1 - p_{202} + \hat{\tau})(1 - p_{303} + \hat{\tau}) + \alpha_3(1 - \phi)p_{103}(1 - p_{202} + \hat{\tau})}{(1 - p_{101} + \hat{\tau})(1 - p_{202} + \hat{\tau})(1 - p_{303} + \hat{\tau})}\\ &\quad + \frac{\alpha_2(1 - \phi)p_{102}(1 - p_{303} + \hat{\tau}) + \alpha_2(1 - \phi)p_{302}p_{103}}{(1 - p_{101} + \hat{\tau})(1 - p_{202} + \hat{\tau})(1 - p_{303} + \hat{\tau})}.\end{aligned}$$

Substituting the above expressions into eqn (5.40) and simplifying eventually gives us

$$\frac{d\mathcal{L}}{dt} = \gamma\pi_1(R_0 - 1) + A^- + B^- + C^-, \tag{5.41}$$

therefore $d\mathcal{L}(\mathbf{x})/dt$ is always (at least) negative semidefinite for $R_0 \leq 1$. This concludes the proof of Lemma 5.1.•

Lemma 5.2 *When $R_0 \leq 1$ the only invariant set in $d\mathcal{L}/dt = 0$ is* $(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3) = (0, 0, 0, 0, 0, 0)$.

Proof.

For a set to be in $d\mathcal{L}/dt = 0$ when $R_0 \leq 1$ we require that A^-, B^- and C^- are all equal to zero in this set. It is obvious that the only set which will give $C^- = 0$ is

the set where $\pi = 0$ or $\beta = 0$. However if $\pi = 0$ and $\beta > 0$ then $d\pi/dt > 0$, and if $\pi > 0$ and $\beta = 0$ then $d\beta/dt > 0$. Hence the only invariant set is (π, β)=(0,0), the disease-free equilibrium solution.•

Lemma 5.3 *If* $R_0 \leq 1$ *and* $p_{011} \geq \max(p_{211}, p_{311})$, $p_{011} + p_{013} \geq p_{211} + p_{213}$ *and* $p_{033} \geq p_{233}$ *then whatever the initial state the disease will die out in both addicts and needles.*

Proof.

From Lemma 5.1 we have that when $R_0 \leq 1$, $\mathcal{L}(\mathbf{x})$ is a weak Lyapunov function for the disease-free solution of eqns (5.17)-(5.22). When $R_0 \leq 1$ we always have that $d\mathcal{L}/dt \leq 0$. In the case where $d\mathcal{L}/dt < 0$ for all $\mathbf{x} > 0$ we have that $\mathcal{L}(\mathbf{x})$ is a strong Lyapunov function for the disease-free solution and hence by Theorem 10.2 in Jordan and Smith (1987) the disease-free solution is globally asymptotically stable. In the case where $d\mathcal{L}/dt = 0$ we have from Lemma 5.2 that the only invariant set in $d\mathcal{L}/dt = 0$ is the disease-free solution. By LaSalle's Invariance Principle, LaSalle (1976), $\mathbf{x}(t) \to M \cap \mathcal{L}^{-1}(c)$ for some $c \geq 0$ where $\mathbf{x} = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_{1+3}, \beta)$ and where M is the largest invariant set in $d\mathcal{L}/dt = 0$. Hence the disease-free solution is again globally asymptotically stable.•

We have now proved our assertions in Theorem 5.1 for the case where $R_0 \leq 1$ we now examine the properties of our model when $R_0 > 1$. Firstly we show that the disease-free equilibrium is no longer stable.

Lemma 5.4 *The disease-free equilibrium is unstable when* $R_0 > 1$.

Proof.

Consider the linearised system of eqns (5.17)-(5.22), evaluated at the disease-free equilibrium. This system can be represented in matrix form as $d\mathbf{x}/dt = \mathbf{J}\mathbf{x}$, where $\mathbf{x}^T = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$ and $\mathbf{J} =$

$$\begin{bmatrix}
-(\mu+\delta_1) & 0 & 0 & \lambda\alpha_1(1-\phi) & \lambda\alpha_2(1-\phi) & \lambda\alpha_3(1-\phi) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\
\lambda\gamma p_{011} & 0 & 0 & -\lambda\gamma(1-p_{101}+\hat{\tau}) & 0 & 0 \\
\lambda\gamma p_{012} & \lambda\gamma p_{022} & \lambda\gamma p_{032} & \lambda\gamma p_{102} & -\lambda\gamma(1-p_{202}+\hat{\tau}) & \lambda\gamma p_{302} \\
\lambda\gamma p_{013} & 0 & \lambda\gamma p_{033} & \lambda\gamma p_{103} & 0 & -\lambda\gamma(1-p_{303}+\hat{\tau})
\end{bmatrix}.$$

We wish to show that at least one eigenvalue of J has a strictly positive real part. Using the Routh-Hurwitz conditions it is sufficient to show that the constant term in the characteristic equation of J, $\omega^6 + a_1\omega^5 + a_2\omega^4 + a_3\omega^3 + a_4\omega^2 + a_5\omega + a_6 = 0$ is strictly negative. It is straightforward to show that $a_6 =$

$$(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)(1-p_{101}+\hat{\tau})\lambda^3\gamma^3(1-p_{202}+\hat{\tau})(1-p_{303}+\hat{\tau})$$

$$-\lambda\alpha_1 p_{011}(1-\phi)(\mu+\delta_2)(\mu+\delta_3)\lambda^3\gamma^3(1-p_{202}+\hat{\tau})(1-p_{303}+\hat{\tau})$$

$$-\lambda\alpha_2(1-\phi)\delta_1(1-p_{101}+\hat{\tau})\delta_2\lambda^3\gamma^3 p_{032}(1-p_{303}+\hat{\tau})$$

$$-\lambda\alpha_2(1-\phi)\delta_1(1-p_{101}+\hat{\tau})\delta_2 p_{302}p_{033}\lambda^3\gamma^3$$

$$-\lambda\alpha_2(1-\phi)\delta_1(1-p_{101}+\hat{\tau})(\mu+\delta_3)p_{022}(1-p_{303}+\hat{\tau})\lambda^3\gamma^3$$

$$-\lambda\alpha_2(1-\phi)(\mu+\delta_2)(\mu+\delta_3)\lambda^3\gamma^3\Big[p_{011}p_{102}(1-p_{303}+\hat{\tau})+p_{013}p_{302}(1-p_{101}+\hat{\tau})$$

$$+p_{012}(1-p_{101}+\hat{\tau})(1-p_{303}+\hat{\tau})+p_{011}p_{103}p_{302}\Big]$$

$$-\lambda\alpha_3(1-\phi)\delta_1\delta_2(1-p_{101}+\hat{\tau})(1-p_{202}+\hat{\tau})\lambda^3\gamma^3 p_{033}$$

$$-\lambda\alpha_3(1-\phi)(\mu+\delta_2)(\mu+\delta_3)(1-p_{202}+\hat{\tau})\lambda^3\gamma^3\Big[p_{011}p_{013}+p_{013}(1-p_{101}+\hat{\tau})\Big].$$

By dividing each term in the above expression by the first term we get $1 - R_0$. Hence

$$a_6 = (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)\lambda^3\gamma^3(1-p_{101}+\hat{\tau})(1-p_{202}+\hat{\tau})(1-p_{303}+\hat{\tau})(1-R_0),$$

therefore it follows directly that $R_0 > 1$ implies that $a_6 < 0$.•

We now show that $\pi_1^\star > 0$ implies that $\pi_2^\star$, $\pi_3^\star$, $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ are all strictly positive and hence the model has two classes of equilibrium solution, a disease-free solution and solutions where disease is present in each type of infectious addict and each type of infectious needle.

Lemma 5.5 *$\pi_1^\star > 0$ determines unique $\pi_2^\star$, $\pi_3^\star$, $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star > 0$ and moreover $\beta^\star < 1$.*

Proof.

From eqn (5.20), eqns (5.20) and (5.22), and eqns (5.20)-(5.22) we have respectively that:

$$\pi_1^\star p_{011} = \beta_1^\star\Big[\pi_1^\star(p_{011}+1-p_{111})+(1-\pi^\star)(1-p_{101})+\pi_2^\star(1-p_{121})+\pi_3^\star(1-p_{131})$$

$$+\hat{\tau}\Big] + \beta_2^\star \pi_1^\star (p_{011} - p_{211}) + \beta_3^\star \pi_1^\star (p_{011} - p_{311}), \tag{5.42}$$

$$\pi_1^\star (p_{011} + p_{013}) + \pi_3^\star p_{033} = \beta_1^\star \Big[\pi_1^\star (p_{011} + p_{013} + 1 - p_{111} - p_{113}) + \pi_2^\star (1 - p_{121} - p_{123})$$

$$+\pi_3^\star (p_{033} + 1 - p_{131} - p_{133}) + (1 - \pi^\star)(1 - p_{101} - p_{103}) + \hat{\tau}\Big] + \beta_2^\star \Big[\pi_1^\star (p_{011} + p_{013}$$

$$-p_{211} - p_{213}) + \pi_3^\star (p_{033} - p_{233})\Big] + \beta_3^\star \Big[\pi_1^\star (p_{011} + p_{013} + 1 - p_{313} - p_{311})$$

$$+\pi_2^\star (1 - p_{323}) + \pi_3^\star (p_{033} + 1 - p_{333}) + (1 - \pi^\star)(1 - p_{303}) + \hat{\tau}\Big], \tag{5.43}$$

and

$$\pi_1^\star (p_{011} + p_{012} + p_{013}) + \pi_2^\star p_{022} + \pi_3^\star (p_{032} + p_{033}) = \beta_1^\star \Big[\pi_1^\star (p_{011} + p_{012} + p_{013} + p_{110})$$

$$+\pi_2^\star (p_{022} + p_{120}) + \pi_3^\star (p_{032} + p_{033} + p_{130}) + (1 - \pi^\star) p_{100} + \hat{\tau}\Big] + \beta_2^\star \Big[\pi_1^\star (p_{011} + p_{012}$$

$$+p_{013} + p_{210}) + \pi_2^\star (p_{022} + p_{220}) + \pi_3^\star (p_{032} + p_{033} + p_{230}) + (1 - \pi^\star) p_{200} + \hat{\tau}\Big]$$

$$+\beta_3^\star \Big[\pi_1^\star (p_{011} + p_{012} + p_{013} + p_{310}) + \pi_2^\star (p_{022} + p_{320}) + \pi_3^\star (p_{032} + p_{033} + p_{330})$$

$$+(1 - \pi^\star) p_{300} + \hat{\tau}\Big]. \tag{5.44}$$

Since $\pi_1^\star > 0$ we have directly that $\pi_2^\star > 0$ and $\pi_3^\star > 0$ from eqns (5.18) and (5.19). We can consider eqns (5.42)-(5.44) as a linear system of equations where $\pi_1^\star$, $\pi_2^\star$, and $\pi_3^\star$ are known positive constants and since this is a linear system we can solve these equations. We now substitute $\tilde{\beta}_i$ and $\tilde{\pi}_j$ for $\beta_i^\star$ and $\pi_j^\star$ respectively where $\tilde{\beta}_i = \frac{\beta_i^\star}{1-\beta^\star}$ and $\tilde{\pi}_j = \frac{\pi_j^\star}{1-\pi^\star}$ for $i, j = 1, 2, 3$. Note that as $d\beta/dt < 0$ when $\beta = 1$ and $d\pi/dt < 0$ when $\pi = 1$ and $\pi_1^\star > 0$ we cannot have that $\beta^\star = 1$ or $\pi^\star = 1$, so we can divide by $1 - \beta^\star$ and $1 - \pi^\star$. This gives us the following system of equations in $\tilde{\beta}_i$ and $\tilde{\pi}_j$ for $i, j = 1, 2, 3$.

$$\tilde{\pi}_1 p_{011} = \tilde{\beta}_1 \Big[1 - p_{101} + \tilde{\pi}_1 (1 - p_{111}) + \tilde{\pi}_2 (1 - p_{121}) + \tilde{\pi}_3 (1 - p_{131}) + \tilde{\tau}\Big] - \tilde{\beta}_2 \tilde{\pi}_1 p_{211} - \tilde{\beta}_3 \tilde{\pi}_1 p_{311}, \tag{5.45}$$

$$\tilde{\pi}_1 (p_{011} + p_{013}) + \tilde{\pi}_3 p_{033} = \tilde{\beta}_1 \Big[1 - p_{101} - p_{103} + \tilde{\pi}_1 (1 - p_{111} - p_{113})$$

$$+\tilde{\pi}_2 (1 - p_{121} - p_{123}) + \tilde{\pi}_3 (1 - p_{131} - p_{133}) + \tilde{\tau}\Big] - \tilde{\beta}_2 \Big[\tilde{\pi}_3 p_{233} + \tilde{\pi}_1 (p_{211} + p_{213})\Big]$$

$$+\tilde{\beta}_3 \Big[1 - p_{303} + \tilde{\pi}_1 (1 - p_{313} - p_{311}) + \tilde{\pi}_2 (1 - p_{323}) + \tilde{\pi}_3 (1 - p_{333}) + \tilde{\tau}\Big], \tag{5.46}$$

and

$$\tilde{\pi}_1 (p_{011} + p_{012} + p_{013}) + \tilde{\pi}_2 p_{022} + \tilde{\pi}_3 (p_{032} + p_{033}) = \tilde{\beta}_1 (p_{100} + \tilde{\pi}_1 p_{110} + \tilde{\pi}_2 p_{120}$$

$$+\tilde{\pi}_3 p_{130} + \tilde{\tau}) + \tilde{\beta}_2(p_{200} + \tilde{\pi}_1 p_{210} + \tilde{\pi}_2 p_{220} + \tilde{\pi}_3 p_{230} + \tilde{\tau}) + \tilde{\beta}_3(p_{300} + \tilde{\pi}_1 p_{310} +$$

$$\tilde{\pi}_2 p_{320} + \tilde{\pi}_3 p_{330} + \tilde{\tau}), \tag{5.47}$$

where $\tilde{\tau} = \hat{\tau}/(1 - \pi^*)$. We now solve this system for $\tilde{\beta}_1$, $\tilde{\beta}_2$ and $\tilde{\beta}_3$ in terms of $\tilde{\pi}_1$, $\tilde{\pi}_2$ and $\tilde{\pi}_3$. We use eqn (5.45) to write $\tilde{\beta}_1$ in terms of $\tilde{\beta}_2$ and $\tilde{\beta}_3$. We then substitute this expression for $\tilde{\beta}_1$ into eqn (5.46) which allows us to write $\tilde{\beta}_3$ in terms of $\tilde{\beta}_2$ only. Hence we have $\tilde{\beta}_1$ and $\tilde{\beta}_3$ in terms of $\tilde{\beta}_2$ only and then use eqn (5.47) to get an explicit expression for $\tilde{\beta}_2$. We have that

$$\tilde{\beta}_1 = \frac{\tilde{\pi}_1(p_{011} + \tilde{\beta}_2 p_{211} + \tilde{\beta}_3 p_{311})}{1 - p_{101} + \tilde{\pi}_1(1 - p_{111}) + \tilde{\pi}_2(1 - p_{121}) + \tilde{\pi}_3(1 - p_{131}) + \tilde{\tau}}. \tag{5.48}$$

We now use eqn (5.48) to substitute for $\tilde{\beta}_1$ in eqn (5.46), this gives us

$$\tilde{\beta}_3 = \frac{1}{A_2}\left[\tilde{\pi}_1\left(p_{011}(p_{103} + \tilde{\pi}_1 p_{113} + \tilde{\pi}_2 p_{123} + \tilde{\pi}_3 p_{133}) + A_1 p_{013}\right) + \tilde{\pi}_3 A_1 p_{033} + \tilde{\beta}_2 A_3\right], \tag{5.49}$$

where

$$A_1 = 1 - p_{101} + \tilde{\pi}_1(1 - p_{111}) + \tilde{\pi}_2(1 - p_{121}) + \tilde{\pi}_3(1 - p_{131}) + \tilde{\tau},$$

$$A_2 = \left[1 - p_{303} + \tilde{\tau} + \tilde{\pi}_1(1 - p_{313} - p_{311}) + \tilde{\pi}_2(1 - p_{323}) + \tilde{\pi}_3(1 - p_{333})\right] A_1$$

$$+\tilde{\pi}_1 p_{311}\left[1 - p_{101} - p_{103} + \tilde{\pi}_1(p_{110} + p_{112}) + \tilde{\pi}_2(p_{122} + p_{120}) + \tilde{\pi}_3(p_{132} + p_{130}) + \tilde{\tau}\right],$$

and

$$A_3 = \tilde{\pi}_1 p_{211}\left[p_{103} + \tilde{\pi}_1 p_{113} + \tilde{\pi}_2 p_{123} + \tilde{\pi}_3 p_{133}\right] + A_1\left[\tilde{\pi}_1 p_{213} + \tilde{\pi}_3 p_{233}\right].$$

Note that A_1 and A_2 are both strictly positive (since $\tilde{\tau} > 0$ always). We now have expressions for $\tilde{\beta}_1$ in terms of $\tilde{\beta}_2$ and $\tilde{\beta}_3$ only and $\tilde{\beta}_3$ in terms of $\tilde{\beta}_2$ only. Hence we can substitute these expressions into eqn (5.47) to get an explicit expression for $\tilde{\beta}_2$. After some simplification equation (5.47) becomes

$$\tilde{\beta}_2 = \frac{E}{D}, \qquad \text{where} \tag{5.50}$$

$$D = A_1\left(p_{200} + \tilde{\pi}_1 p_{210} + \tilde{\pi}_2 p_{220} + \tilde{\pi}_3 p_{230} + \tilde{\tau}\right) + \tilde{\pi}_1 p_{211}\left(p_{100} + \tilde{\pi}_1 p_{110} + \tilde{\pi}_2 p_{120}\right.$$

$$\left.+\tilde{\pi}_3 p_{130} + \tilde{\tau}\right) + \left\{A_1\left(p_{300} + \tilde{\pi}_1 p_{310} + \tilde{\pi}_2 p_{320} + \tilde{\pi}_3 p_{330} + \tilde{\tau}\right) + \tilde{\pi}_1 p_{311}\left(p_{100}\right.\right.$$

$$\left.\left.+\tilde{\pi}_1 p_{110} + \tilde{\pi}_2 p_{120} + \tilde{\pi}_3 p_{130} + \tilde{\tau}\right)\right\}\frac{A_3}{A_2}, \tag{5.51}$$

and

$$\begin{aligned} E &= \tilde{\pi}_1\Big(A_1(p_{011}+p_{012}+p_{013}) - p_{011}[p_{100}+\tilde{\pi}_1 p_{110}+\tilde{\pi}_2 p_{120}+\tilde{\pi}_3 p_{130}+\tilde{\tau}]\Big) \\ &\quad + A_1(\tilde{\pi}_2 p_{022}+\tilde{\pi}_3(p_{032}+p_{033})) - \frac{1}{A_2}\Big\{\Big[A_1(p_{300}+\tilde{\pi}_1 p_{310}+\tilde{\pi}_2 p_{320}+\tilde{\pi}_3 p_{330}+\tilde{\tau}) \\ &\quad + \tilde{\pi}_1 p_{311}(p_{100}+\tilde{\pi}_1 p_{110}+\tilde{\pi}_2 p_{120}+\tilde{\pi}_3 p_{130}+\tilde{\tau})\Big] \\ &\quad \times\Big[\tilde{\pi}_1\Big(p_{011}(p_{103}+\tilde{\pi}_1 p_{113}+\tilde{\pi}_2 p_{123}+\tilde{\pi}_3 p_{133})+A_1 p_{013}\Big)+\tilde{\pi}_3 p_{033}A_1\Big]\Big\}. \end{aligned} \tag{5.52}$$

It is straightforward to show that $\tilde{\beta}_2 > 0$ since clearly $D > 0$ and

$$\begin{aligned} &p_{011}A_1 - p_{011}\Big[p_{100}+\tilde{\pi}_1 p_{110}+\tilde{\pi}_2 p_{120}+\tilde{\pi}_3 p_{130}+\tilde{\tau}\Big] \\ &\quad = p_{011}\Big\{p_{102}+p_{103}+\tilde{\pi}_1(p_{112}+p_{113})+\tilde{\pi}_2(p_{122}+p_{123})+\tilde{\pi}_3(p_{133}+p_{132})\Big\}, \\ &\quad > p_{011}\Big(p_{103}+\tilde{\pi}_1 p_{113}+\tilde{\pi}_2 p_{123}+\tilde{\pi}_3 p_{133}\Big), \end{aligned}$$

and

$$\begin{aligned} A_2 \;\geq\; &\Big(p_{300}+\tilde{\pi}_1 p_{310}+\tilde{\pi}_2 p_{320}+\tilde{\pi}_3 p_{330}+\tilde{\tau}\Big)A_1 + \tilde{\pi}_1 p_{311}\Big(p_{100}+\tilde{\pi}_1 p_{110}+\tilde{\pi}_2 p_{120} \\ &+\tilde{\pi}_3 p_{130}+\tilde{\tau}\Big), \end{aligned}$$

from which it follows that $E > 0$. We have that $\tilde{\beta}_2 > 0$, using eqn (5.49) this implies that $\tilde{\beta}_3 > 0$, and using eqn (5.48) implies that $\tilde{\beta}_1 > 0$ also. Hence as $\beta_i^\star = \tilde{\beta}_i/(1+\tilde{\beta})$ for $i = 1,2,3$ and $\beta^\star = \tilde{\beta}/(1+\tilde{\beta})$ we have that $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ are all strictly positive and moreover $\beta^\star < 1$. This concludes the proof of Lemma 5.5.•

We now use the previous lemma to show that when $R_0 > 1$ there exists at least one strictly positive endemic equilibrium solution.

Lemma 5.6 *If $R_0 > 1$ then there exists at least one endemic equilibrium solution to eqns (5.17)-(5.22).*

Proof.

We have now implicitly expressed $\pi_2^\star$, $\pi_3^\star$, $\beta_1^\star$, $\beta_2^\star$ and $\beta_3^\star$ in terms of $\pi_1^\star$. The equation which determines the endemic equilibrium value of $\pi_1^\star$ is $F(\pi_1^\star) = 1$ where

$$F(\pi_1^\star) = \frac{\lambda(1-\phi)}{\mu+\delta_1}\frac{(\beta_1^\star\alpha_1+\beta_2^\star\alpha_2+\beta_3^\star\alpha_3)}{\tilde{\pi}_1}. \tag{5.53}$$

We assert that eqn (5.53) has at least one strictly positive solution. Let $k_2 = \delta_1/(\mu+\delta_2)$ and $k_3 = (\delta_1\delta_2)/((\mu+\delta_2)(\mu+\delta_3))$. We have $0 \leq \pi^\star \leq 1$ so $0 \leq \pi_1^\star \leq \frac{1}{1+k_2+k_3}$. As

$\pi_1^\star \to \frac{1}{1+k_2+k_3}$, $\tilde{\pi}_1 = \frac{\pi_1^\star}{1-\pi^\star} \to \infty$. We have that $0 \leq \beta_1^\star\alpha_1 + \beta_2^\star\alpha_2 + \beta_3^\star\alpha_3 \leq \alpha_1+\alpha_2+\alpha_3$, so $F(\pi_1^\star) \to 0$ as $\pi_1^\star \to \frac{1}{1+k_2+k_3}$. As $\pi_1^\star \to 0$, $\tilde{\pi}_1$, $\tilde{\pi}_2$, $\tilde{\pi}_3 \to 0$ and from the proof of Lemma 5.5 $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3 \to 0$ also. Eqns (5.45)-(5.47) become

$$\tilde{\pi}_1 p_{011} + o(\tilde{\pi}_1) = \tilde{\beta}_1(1 - p_{101} + \tilde{\tau}), \tag{5.54}$$

$$(p_{011} + p_{013} + k_3 p_{033})\tilde{\pi}_1 + o(\tilde{\pi}_1) = \tilde{\beta}_1(1 - p_{101} - p_{103} + \tilde{\tau}) + \tilde{\beta}_3(1 - p_{303} + \tilde{\tau}), \tag{5.55}$$

and

$$(p_{011} + p_{012} + p_{013} + k_2 p_{022} + k_3(p_{032} + p_{033}))\tilde{\pi}_1 + o(\tilde{\pi}_1) = \tilde{\beta}_1(1 - p_{101} - p_{102} - p_{103} + \tilde{\tau}) + \tilde{\beta}_2(1 - p_{202} + \tilde{\tau}) + \tilde{\beta}_3(1 - p_{303} - p_{302} + \tilde{\tau}). \tag{5.56}$$

We have that

$$\lim_{\pi_1^\star \to 0} F(\pi_1^\star) = \lim_{\pi_1^\star \to 0} \frac{\lambda(1-\phi)}{\mu + \delta_1} \frac{(\beta_1^\star\alpha_1 + \beta_2^\star\alpha_2 + \beta_3^\star\alpha_3)}{\tilde{\pi}_1},$$

$$= \lim_{\pi_1^\star \to 0} \frac{\lambda}{\mu + \delta_1} \frac{1}{\tilde{\pi}_1}$$

$$\times \Big[\beta_1^\star(1 - p_{101} + \hat{\tau})\zeta_1 + \Big[\beta_1^\star(1 - p_{101} - p_{103} + \hat{\tau}) + \beta_3^\star(1 - p_{303} + \hat{\tau})\Big]\zeta_2$$

$$+\Big[\beta_1^\star(1 - p_{101} - p_{102} - p_{103} + \hat{\tau}) + \beta_2^\star(1 - p_{202} + \hat{\tau})$$

$$+\beta_3^\star(1 - p_{303} - p_{302} + \hat{\tau})\Big]\zeta_3\Big], \tag{5.57}$$

by construction of ζ_1, ζ_2 and ζ_3. Now $\pi_1^\star \to 0 \Leftrightarrow \tilde{\pi}_1 \to 0$, $\beta_i^\star = \tilde{\beta}_i/(1+\tilde{\beta}) = \tilde{\beta}_i + o(\tilde{\pi}_1)$ for $i = 1,2,3$ and $\hat{\tau} = \tilde{\tau}/(1-\pi^\star) = \tilde{\tau} + o(\tilde{\pi}_1)$. Therefore $\lim_{\pi_1^\star \to 0} F(\pi_1^\star)$

$$= \lim_{\pi_1^\star \to 0} \frac{\lambda}{\mu + \delta_1} \frac{1}{\tilde{\pi}_1}$$

$$\times \Big[\tilde{\beta}_1(1 - p_{101} + \tilde{\tau})\zeta_1 + \Big[\tilde{\beta}_1(1 - p_{101} - p_{103} + \tilde{\tau}) + \tilde{\beta}_3(1 - p_{303} + \tilde{\tau})\Big]\zeta_2$$

$$+\Big[\tilde{\beta}_1(p_{100} + \tilde{\tau}) + \tilde{\beta}_2(p_{200} + \tilde{\tau}) + \tilde{\beta}_3(p_{300} + \tilde{\tau})\Big]\zeta_3 + o(\tilde{\pi}_1)\Big], \tag{5.58}$$

$$= \frac{\lambda}{\mu + \delta_1}\Big[\zeta_1 p_{011} + \zeta_2(p_{011} + p_{013} + k_3 p_{033})$$

$$+\zeta_3(p_{011} + p_{012} + p_{013} + k_2 p_{022} + k_3(p_{032} + p_{033}))\Big] \tag{5.59}$$

by eqns (5.54)-(5.56),

$$= \quad R_0 \qquad \text{(after some simplification).} \tag{5.60}$$

Hence $F(\pi_1^\star) \to 0$ as $\pi_1^\star \to \frac{1}{(1+k_2+k_3)}$ and $F(\pi_1^\star) \to R_0$ as $\pi_1^\star \to 0$. So if $R_0 > 1$ then $F(\pi_1^\star) = 1$ has at least one root in $\left(0, \frac{1}{(1+k_2+k_3)}\right)$. The proof of Lemma 5.5 shows that each value of $\pi_1^\star$ in $\left(0, \frac{1}{(1+k_2+k_3)}\right)$ corresponds to a unique feasible endemic equilibrium solution of eqns (5.17)-(5.22).•

Lemma 5.7 *If $R_0 > 1$ and either $\pi(0) > 0$ or $\beta(0) > 0$ then there exists a fixed $\epsilon > 0$ depending only on the model parameters and not the initial conditions such that for some $T > 0$, $\pi_i(t) \geq \epsilon$ and $\beta_j(t) \geq \epsilon$ for $i, j = 1, 2, 3$, for all $t \geq T(\epsilon)$.*

Proof.

Proving the persistence of disease when $R_0 > 1$ in the General Model follows the same intuitive method as the equivalent results for the Simple, Optimistic and Pessimistic Models. Firstly using a method similar to Lemma 4.1 it is straightforward to show that if initially any of $\pi_i(0)$ or $\beta_i(0)$ for $i = 1, 2, 3$ are strictly positive then $\pi_i(\Delta t) > 0$ and $\beta_i(\Delta t) > 0$ for $i = 1, 2, 3$ and Δt small and strictly positive. Hence if disease is initially present then $\pi_1(\Delta t) > \epsilon\pi_1^\star$ for ϵ small enough. We now show that once π_1 has increased to this level it can be bounded away from the origin for all $t \geq \Delta t$. Again following a similar method to previously we first show that if $\pi_1(t)$ remains continuously below $\frac{1}{2}\epsilon\pi_1^\star$ then all other model components also become small.

We can use Lemma 2.4 and Corollary 2.5 directly to bound above π_2 and π_3 when $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$. We also wish to show that β_i for $i = 1, 2, 3$ can be bounded above when $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$. By examining the General Model equations it is easy to see that

$$\frac{d\beta}{dt} \quad \leq \quad \lambda\gamma\pi - \beta\tau.$$

Hence using method of Lemma 2.4 and the results of Lemma 2.4 and Corollary 2.5 we have that in $[t_0, t_1]$

$$\frac{d\beta}{dt} \quad \leq \quad \lambda\gamma\left(\frac{1}{2} + 2\Delta\right)\epsilon\pi^* - \beta\tau, \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_2.$$

Integrating over $[t_0 + \bar{T}_1 + \bar{T}_2, t]$ gives us

$$\beta(t) \quad \leq \quad \left(\frac{1}{2} + 3\Delta\right)\epsilon_s\beta^\star, \qquad \text{for } t \geq t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3,$$

where $\epsilon_s = (\lambda\gamma\epsilon\pi^\star)/(\tau\beta^\star)$ is a small strictly positive fixed value and $\bar{T}_3$ is sufficiently large. By bounding above $\beta(t)$ we have also bounded above $\beta_i(t)$ for $i = 1, 2, 3$. Therefore for $t \geq t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3$ we have that $\beta_i(t) \leq \left(\frac{1}{2} + 3\Delta\right)\epsilon_i\beta_i^\star$ for $i = 1, 2, 3$ where $\epsilon_i = \epsilon_s(\beta^\star/\beta_i^\star)$ are small strictly positive fixed values.

Following the method of Lemma 4.2 we have that for $t_1 > t > t_0 + \bar{T}_1 + \bar{T}_2$, $\pi < \sigma$ where σ is small and strictly positive and without loss of generality we assume that $\sigma < 1$. Using eqns (5.17)-(5.22),

$$\frac{d\pi_1}{dt} \geq (1-\sigma)\lambda(1-\phi)(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3) - (\mu+\delta_1)\pi_1, \tag{5.61}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2, \tag{5.62}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3, \tag{5.63}$$

$$\begin{aligned}\frac{d\beta_1}{dt} \geq\ & \lambda\gamma\pi_1 p_{011} - \lambda\gamma\beta_1\sigma p_{011} - \lambda\gamma\beta_1(1-p_{101}) - \lambda\gamma\beta_1\sigma(1-p_{121}) \\ & -\lambda\gamma\beta_1\sigma(1-p_{131}) - \lambda\gamma\beta_1\sigma(1-p_{111}) - \beta_1\tau,\end{aligned} \tag{5.64}$$

$$\begin{aligned}\frac{d\beta_2}{dt} \geq\ & \lambda\gamma\pi_2 p_{022} - \lambda\gamma\beta\sigma p_{022} - \lambda\gamma\beta_2(1-p_{202}) - \lambda\gamma\beta_2\sigma(1-p_{212}) \\ & -\lambda\gamma\beta_2\sigma(1-p_{232}) - \lambda\gamma\beta_2\sigma(1-p_{222}) + \lambda\gamma\beta_1 p_{102} - \lambda\gamma\sigma\beta_1 p_{102} \\ & +\lambda\gamma\beta_3 p_{302} - \lambda\gamma\beta_3\sigma p_{302} + \lambda\gamma\pi_1 p_{012} - \lambda\gamma\sigma\beta p_{012} + \lambda\gamma\pi_3 p_{032} \\ & -\lambda\gamma\sigma\beta p_{032} - \beta_2\tau,\end{aligned} \tag{5.65}$$

and

$$\begin{aligned}\frac{d\beta_3}{dt} \geq\ & \lambda\gamma\pi_3 p_{033} - \lambda\gamma\beta\sigma p_{033} - \lambda\gamma\beta_3(1-p_{303}) - \lambda\gamma\beta_3\sigma(1-p_{313}) \\ & -\lambda\gamma\beta_3\sigma(1-p_{323}) - \lambda\gamma\beta_3\sigma(1-p_{333}) + \lambda\gamma\pi_1 p_{013} - \lambda\gamma\sigma\beta p_{013} \\ & +\lambda\beta_1 p_{103} - \lambda\beta_1\sigma p_{103} - \beta_3\tau.\end{aligned} \tag{5.66}$$

When $\sigma = 0$ the equations obtained by treating the inequalities in (5.61)-(5.66) as equalities represent the linearised form of the General Model evaluated at the disease-free equilibrium. As in Lemma 4.2 these equations can be written in matrix form as $d\mathbf{x}/dt = \mathbf{J}(\sigma)\mathbf{x}$ where $\mathbf{x} = (\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3)$ and $\mathbf{J}(0)$ is the linearised stability matrix about the disease-free equilibrium given in Lemma 5.4. Hence following the same method as in Lemma 4.2 and using t_2 and $\mathbf{e}$ as defined there we have that after a time $t_0 + t_2 + \bar{T}_4$,

$$\begin{aligned}\mathbf{e}.\mathbf{x}(t) > \mathbf{e}\Big(&\frac{1}{2}\epsilon\pi_1^\star, \Big(\frac{1}{2}+\Delta\Big)\epsilon\pi_2^\star, \Big(\frac{1}{2}+2\Delta\Big)\epsilon\pi_3^\star, \Big(\frac{1}{2}+3\Delta\Big)\epsilon_1\beta_1^\star, \Big(\frac{1}{2}+3\Delta\Big)\epsilon_2\beta_2^\star, \\ &\Big(\frac{1}{2}+3\Delta\Big)\epsilon_3\beta_3^\star\Big),\end{aligned}$$

where $\bar{T}_4$ depends only on ϵ, Δ and the model parameters. From our previous results already know that provided that $t_0 \leq t \leq t_1$ then after a time $t_0 + \bar{T}_1 + \bar{T}_2 + \bar{T}_3$,

$$\mathbf{e.x}(t) \leq \mathbf{e}\left(\frac{1}{2}\epsilon\pi_1^\star, \left(\frac{1}{2}+\Delta\right)\epsilon\pi_2^\star, \left(\frac{1}{2}+2\Delta\right)\epsilon\pi_3^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_1\beta_1^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_2\beta_2^\star, \left(\frac{1}{2}+3\Delta\right)\epsilon_3\beta_3^\star\right).$$

Therefore if $t > t_0 + \max[\bar{T}_1 + \bar{T}_2 + \bar{T}_3, t_2 + \bar{T}_4]$ then we have a contradiction, from which it follows that π_1 only remains continuously below the level $\frac{1}{2}\epsilon\pi_1^\star$ for at most a (fixed and finite) duration T which depends only on ϵ, Δ and the model parameters. Using a similar argument to that in the corresponding result for the Pessimistic Model we deduce that for all $t \geq \Delta t$, $\pi_1(t) \geq \frac{1}{2}\epsilon\pi_1^\star \exp[-(\mu+\delta_1)T]$, which is strictly positive.

We have shown that if disease is initially present in the General Model then there will always be some addicts in stage one infectivity present in the population. Since for $t \geq \Delta t$, $\pi_1 > \epsilon^+$ for ϵ^+ fixed and strictly positive then $\pi_{1,\infty} > 0$, and Lemma 2.2 and Corollary 2.3 imply that $\pi_{2,\infty} > 0$ and $\pi_{3,\infty} > 0$ so that at least for all $t \geq \eta$, for some sufficiently large $\eta > 0$, there will always be some addicts in infectious stages two and three present in the population. From eqn (5.20),

$$\frac{d\beta_1}{dt} \geq \lambda\gamma\pi_1 p_{011} - \lambda\gamma\beta_1\left[p_{011} + 1 - p_{101} + 1 - p_{121} + 1 - p_{131} + 1 - p_{111} + \hat{\tau}\right],$$

therefore using the method of Lemma 2.2 we have directly that

$$\beta_{1,\infty} \geq \frac{p_{011}\pi_{1,\infty}}{p_{011} + 1 - p_{101} + 1 - p_{121} + 1 - p_{131} + 1 - p_{111} + \hat{\tau}}.$$

Hence we also have that if η is sufficiently large then for all $t \geq \eta$, $\beta_1 > \epsilon_1^+$ for some strictly positive fixed ϵ_1^+, so there will always be some needles which are in infectious state one.

We have that

$$\begin{aligned}\frac{d\beta}{dt} &\leq \lambda\gamma\left[\pi_1(p_{011}+p_{012}+p_{013}) + \pi_2 p_{022} + \pi_3(p_{032}+p_{033})\right](1-\beta) - \beta\tau, \\ &\leq \lambda\gamma(1-\beta) - \beta\tau.\end{aligned}$$

Hence writing $\Psi = 1 - \beta$,

$$\begin{aligned}\frac{d\Psi}{dt} &\geq -\lambda\gamma\Psi + \tau(1-\Psi), \\ &= \tau - (\tau + \lambda\gamma)\Psi.\end{aligned}$$

It is straightforward to show that $\Psi^{\infty} \geq \frac{\tau}{\tau+\lambda\gamma} > 0$. Hence provided that $t \geq \eta$ for η sufficiently large $1-\beta \geq \epsilon_0^+ > 0$ where $\epsilon_0^+ = \frac{1}{2}\frac{1}{1+\hat{\tau}} > 0$. We can now use this result to bound $\beta_{2,\infty}$ and $\beta_{3,\infty}$ below. For $t \geq \eta$,

$$\frac{d\beta_3}{dt} \geq \lambda\gamma\pi_3 p_{033}\epsilon_0^+ - \lambda\gamma\beta_3\Big[1-p_{303}+1-p_{313}+1-p_{323}+1-p_{333}+\hat{\tau}\Big],$$

and

$$\frac{d\beta_2}{dt} \geq \lambda\gamma\pi_2 p_{022}\epsilon_0^+ - \lambda\gamma\beta_2\Big[1-p_{202}+1-p_{212}+1-p_{232}+1-p_{222}+\hat{\tau}\Big].$$

Therefore

$$\beta_{2,\infty} \geq \frac{p_{022}\epsilon_0^+\pi_{2,\infty}}{1-p_{202}+1-p_{212}+1-p_{232}+1-p_{222}+\hat{\tau}}$$

and

$$\beta_{3,\infty} \geq \frac{p_{033}\epsilon_0^+\pi_{3,\infty}}{1-p_{303}+1-p_{313}+1-p_{323}+1-p_{333}+\hat{\tau}}.$$

So we additionally have that there will always be some needles which are in infectious states two and three (at least for all $t \geq \eta_1$ for η_1 sufficiently large). This completes the proof.•

This completes our analytical results. Unfortunately we were not able to show analytically, uniqueness or any local or global stability results for the endemic equilibrium when $R_0 > 1$, or produce specific counter-examples where the endemic equilibrium was not unique or not locally stable. A local stability analysis for the endemic equilibrium would involve looking at the roots of a sixth order polynomial so we did not attempt it. It would be theoretically possible to attempt a local stability analysis of a reduced "addict-only" model along the lines of the corresponding proofs for the Optimistic and Pessimistic models. This would involve looking at the roots of only a third order polynomial. However the algebra involved appears horrendous so we did not attempt it.

5.7 Simulation Study of the General Mixing Model

We now use simulation to examine the long term behaviour of the General Model. Firstly we should like to verify that when $R_0 \leq 1$ disease is eventually eradicated from the population irrespective of the initial conditions of the system and individual

parameter estimates. Secondly we wish to investigate whether the prevalence of disease tends to an endemic equilibrium solution when $R_0 > 1$, and disease is initially present in at least one addict or needle, for addict-needle interaction assumptions which are less extreme than those in the Optimistic and Pessimistic Models.

5.7.1 Parameter Estimates

The main reason for constructing the General Model is to examine the effects of different addict-needle interaction assumptions on the spread of HIV. Hence we are more interested in the behaviour of the General Model for different values of the $p^{\star}_{ijk}$ terms than for different values of the other model parameters. In addition whilst the parameter estimates in Table 3.1 are obviously subject to a certain amount of error we do believe that they are reasonably realistic. On the other hand we have very little knowledge concerning likely values of the $p^{\star}_{ijk}$ terms, while we do believe that inequalities (5.1)-(5.10) should always be satisfied this still leaves a very large amount of uncertainty regarding the estimation of these parameters. Hence it seems more important to keep the non $p^{\star}_{ijk}$ parameter estimates fixed at the values estimated previously and vary the $p^{\star}_{ijk}$ parameters in our simulations. This has the added advantage of facilitating a direct comparison of our numerical results with those obtained earlier for the Optimistic and Pessimistic Models.

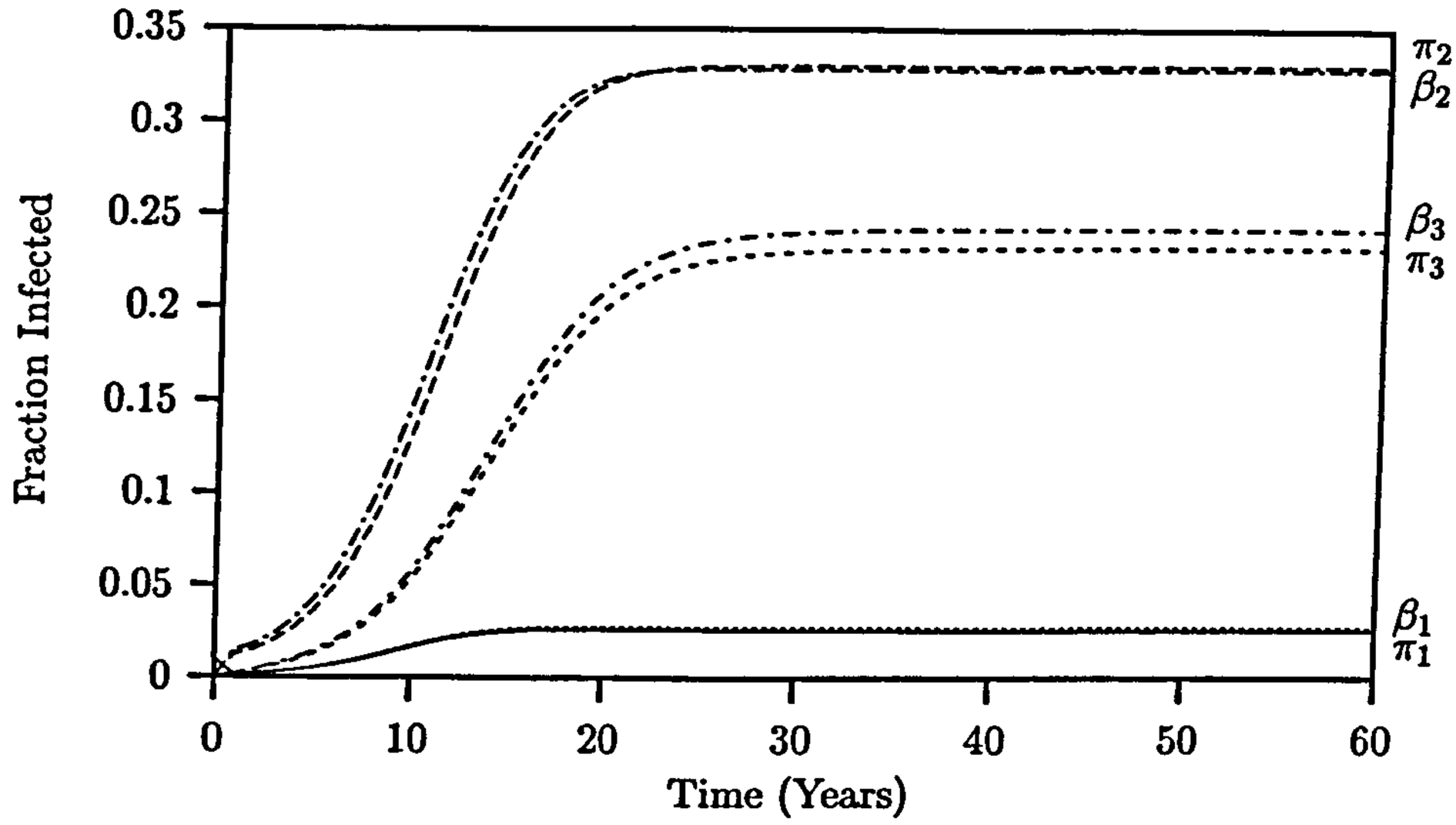
Table 5.4 shows the sets of addict-needle interaction assumptions which we shall use in simulations of the Restricted General Model. It is important to note that the $p^{\star}_{ijk}$ values in this table do not include cleaning. The reason for this is that it is easier to assess how reasonable any particular choice of $p^{\star}_{ijk}$ terms are by considering "blood mixing only" rather than complicating matters by including cleaning. As discussed previously it is easy to adjust $p^{\star}_{ijk}$ terms based on "blood mixing only" to incorporate cleaning. We look at the numerical implications of addicts cleaning needles prior to use in a later chapter.

The second and last columns in Table 5.4 contain the addict-needle interaction assumptions corresponding to the Optimistic and Pessimistic Models respectively. The addict-needle interaction assumptions in columns 3-5 are less extreme. In moving from column 3 to column 5 we have tried to choose addict-needle interaction assumptions which progress from the infectious stage of the addict being most influential (similar to full flushing in single stage models) to the infectious state of the needle being most

Table 5.4: Addict-Needle Assumptions (No Cleaning)

$p^{\star}_{ijk}$	Optim	A	B	C	Pessim
$p^{\star}_{010}$	0	0	0	0	0
$p^{\star}_{011}$	1	1	1	1	1
$p^{\star}_{012}$	0	0	0	0	0
$p^{\star}_{013}$	0	0	0	0	0
$p^{\star}_{020}$	0	0	0	0	0
$p^{\star}_{022}$	1	1	1	1	1
$p^{\star}_{030}$	0	0	0	0	0
$p^{\star}_{032}$	0	0	0	0	0
$p^{\star}_{033}$	1	1	1	1	1
$p^{\star}_{100}$	1	0.7	0.25	0.05	0
$p^{\star}_{101}$	0	0.05	0.25	0.7	1
$p^{\star}_{102}$	0	0.2	0.25	0.05	0
$p^{\star}_{103}$	0	0.05	0.25	0.2	0
$p^{\star}_{121}$	0	0.1	0.33	0.7	1
$p^{\star}_{122}$	1	0.7	0.33	0.1	0
$p^{\star}_{123}$	0	0.2	0.33	0.2	0
$p^{\star}_{131}$	0	0.3	0.5	0.7	1
$p^{\star}_{133}$	1	0.7	0.5	0.3	0
$p^{\star}_{200}$	1	0.7	0.5	0.3	0
$p^{\star}_{202}$	0	0.3	0.5	0.7	1
$p^{\star}_{211}$	1	1	1	1	1
$p^{\star}_{212}$	0	0	0	0	0
$p^{\star}_{213}$	0	0	0	0	0
$p^{\star}_{232}$	0	0	0	0	0
$p^{\star}_{233}$	1	1	1	1	1
$p^{\star}_{300}$	1	0.7	0.33	0.1	0
$p^{\star}_{302}$	0	0.2	0.33	0.2	0
$p^{\star}_{303}$	0	0.1	0.33	0.7	1
$p^{\star}_{311}$	1	1	1	1	1
$p^{\star}_{313}$	0	0	0	0	0
$p^{\star}_{322}$	1	0.7	0.5	0.3	0
$p^{\star}_{323}$	0	0.3	0.5	0.7	1

Figure 5.1: Restricted General Mixing Model (Individual Components)



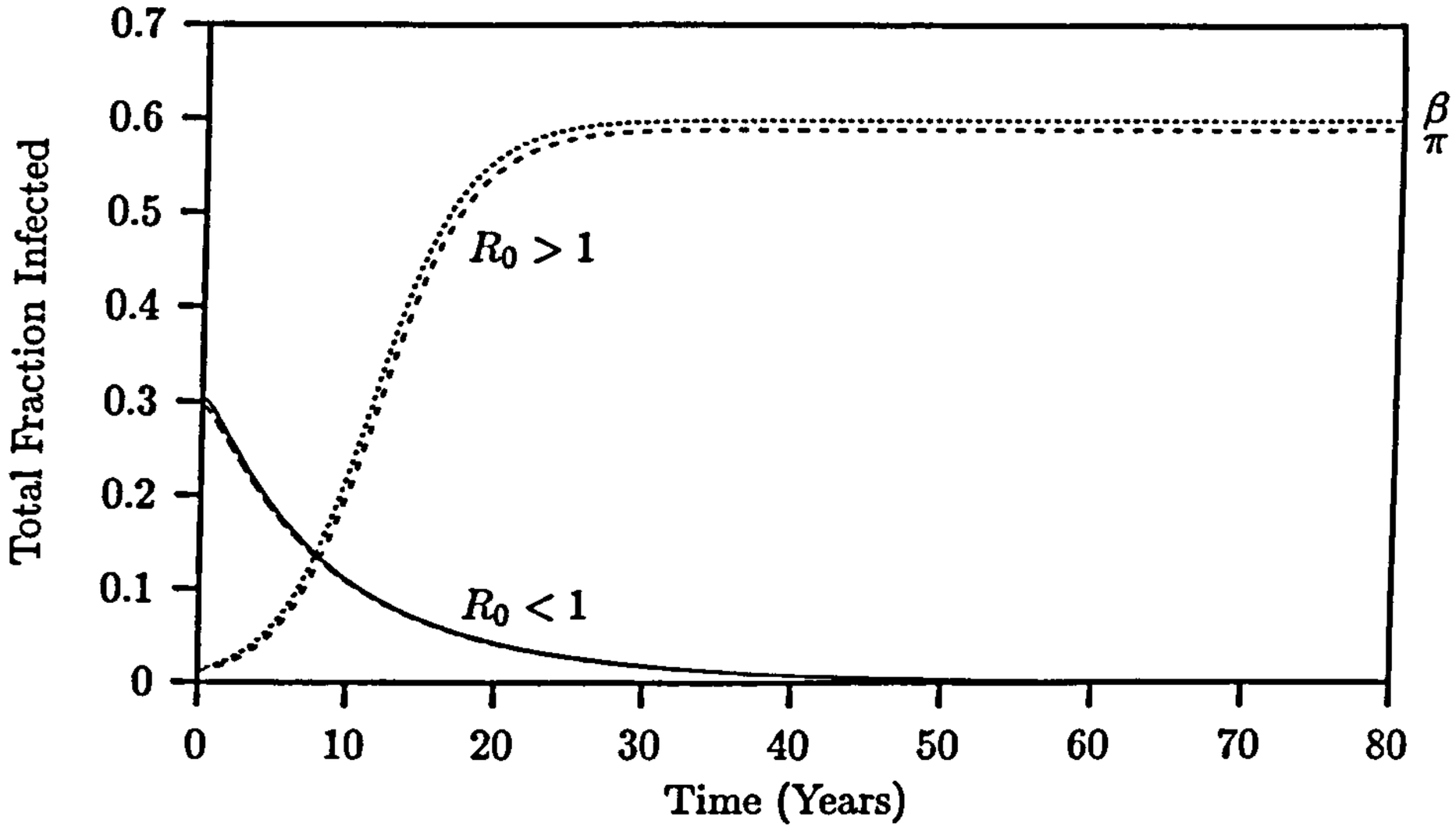
influential (similar to no flushing in single stage models). All five sets of $p^{\star}_{ijk}$ values in Table 5.4 satisfy the inequalities (5.1)-(5.10).

By inspecting the various columns in Table 5.4 it should be apparent that the Optimistic and Pessimistic Models have the most extreme addict-needles interaction assumptions possible. As already mentioned this is due to the assumption that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$. Using the inequalities in eqns (5.1) and (5.7) this additionally implies that $p^{\star}_{211} = 1$, $p^{\star}_{233} = 1$ and $p^{\star}_{311} = 1$. Hence the only probabilities which we have the freedom to alter (but which still must satisfy the inequalities in eqns (5.1)-(5.10)) are $p^{\star}_{10k}$, $p^{\star}_{12k}$, $p^{\star}_{13k}$, $p^{\star}_{20k}$, $p^{\star}_{30k}$ and $p^{\star}_{32k}$. For each of these events the Optimistic Model assumes that the needle is always left in the least infectious state with probability one and likewise the Pessimistic Model assumes that the needle is always left in the most infectious state with probability one. Hence any other choice of addict-needle interaction assumptions must lie between these extremes and we therefore expect the long term prevalence of HIV (in either needles or addicts) in a realistic model to lie between that of the Optimistic and Pessimistic Models.

5.7.2 Simulations

We now simulate the model in eqns (5.11)-(5.16) using addict-needle interaction assumption B, these estimates give an R_0 value of 2.6. Figure 5.1 shows the dynamic behaviour of the prevalence of disease in each class of infectious addict and in each class

Figure 5.2: Restricted General Mixing Model (Total Prevalence)



of infectious needle. It was initially assumed that one percent of the total addict population were acutely infectious and no other addicts or needle were infectious. The figure suggests that the model eventually reaches an endemic equilibrium state. Figure 5.2 illustrates the corresponding total prevalence of HIV in addicts over each of the three infectious classes and similarly for needles. Also shown in this figure is a simulation of the model where now $R_0 = 0.85$, this was achieved by increasing the probability of needle cleaning (denoted by ϕ) from 0.64 to 0.9. In the latter simulation it was initially assumed that 30% of the total addict population were infectious and similarly for the needle population and where all infectious addicts and needles were classed as acutely infectious. The figure clearly shows that for this set of parameter estimates the disease eventually dies out in all addicts and all needles.

Figures 5.3 and 5.4 show simulations of the total fraction of infected addicts in the Restricted General Model when $R_0 > 1$ for the addict-needle interaction assumptions Optim, A, C and Pessim. The values of R_0 for these addict-needle interaction assumptions are 2.200, 2.362, 2.948 and 3.317 respectively. Initially a proportion 0.01 of all addicts are infectious in the simulations in Figure 5.3 where all these addicts are in stage one infectivity, no other addicts or needles are initially infected. Initially a proportion 0.9 of all addicts are infectious in the simulations in Figure 5.4 where all these addicts are in stage one infectivity and again no other addicts or needles are initially infected. It is clear from each of these figures that eventually the disease reaches an en-

Figure 5.3: $p^{\star}_{ijk}$ Selections: Optim, A, C, Pessim ($\pi_1(0) = 0.01$)

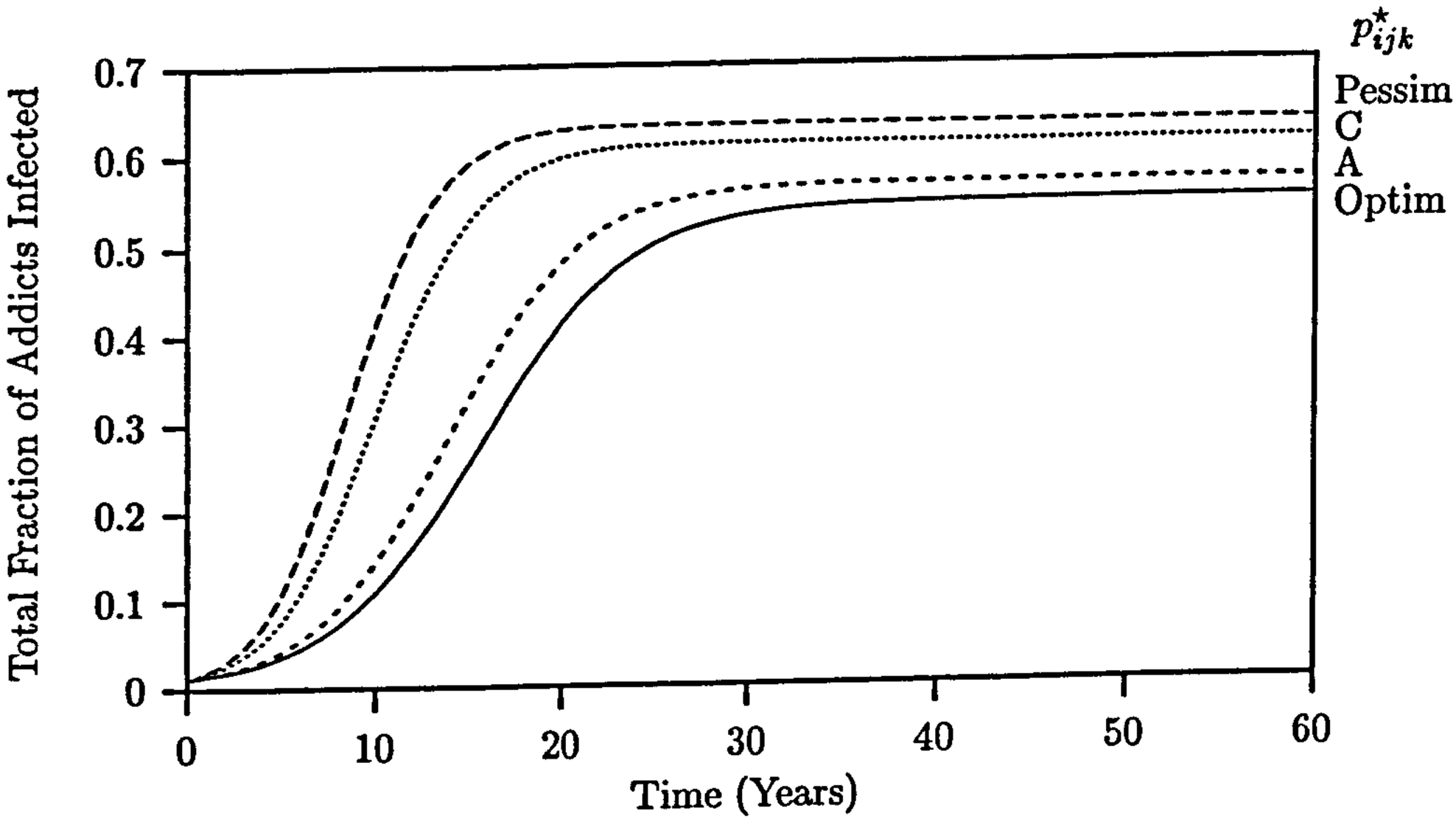


Figure 5.4: $p^{\star}_{ijk}$ Selections: Optim, A, C, Pessim ($\pi_1(0) = 0.9$)

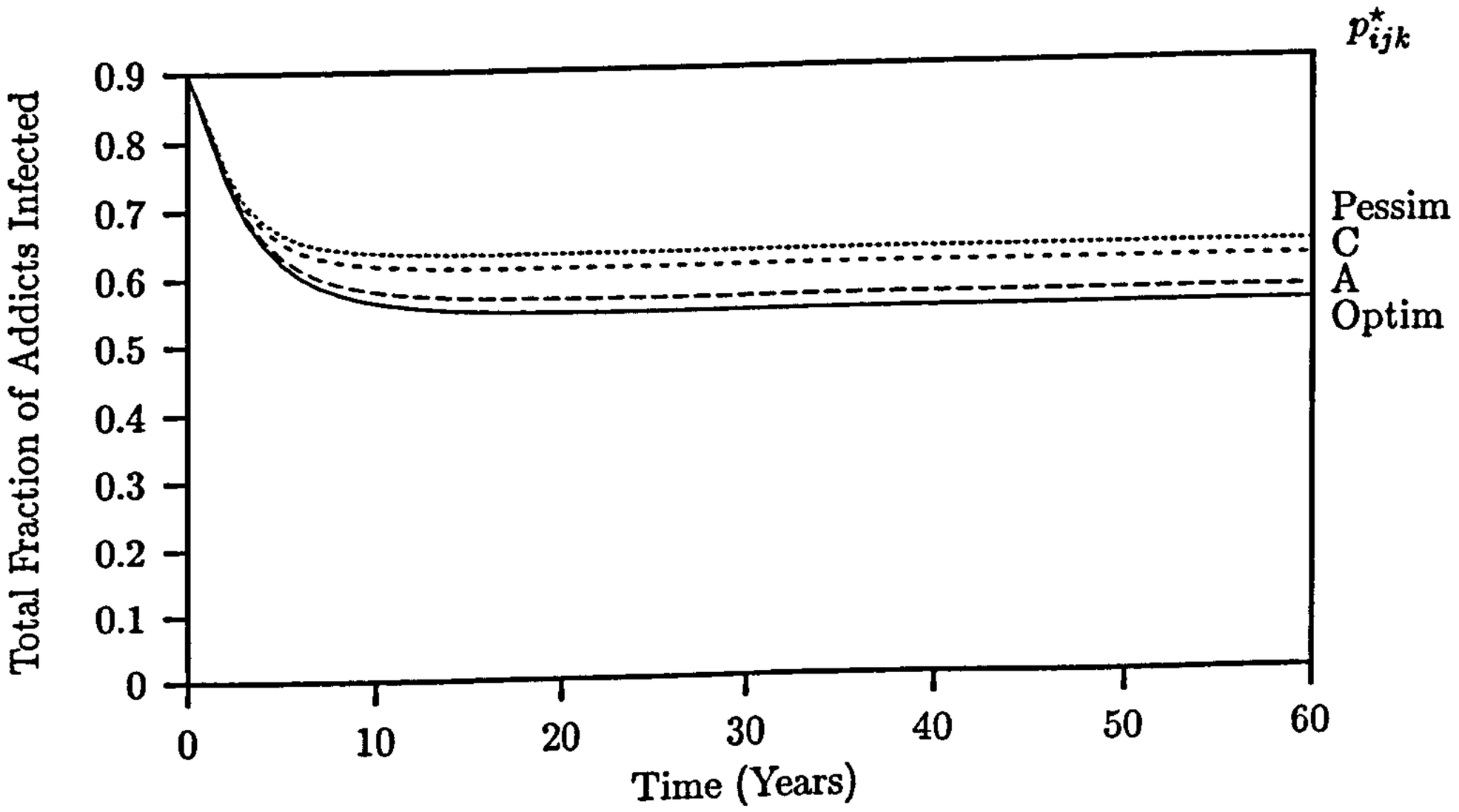
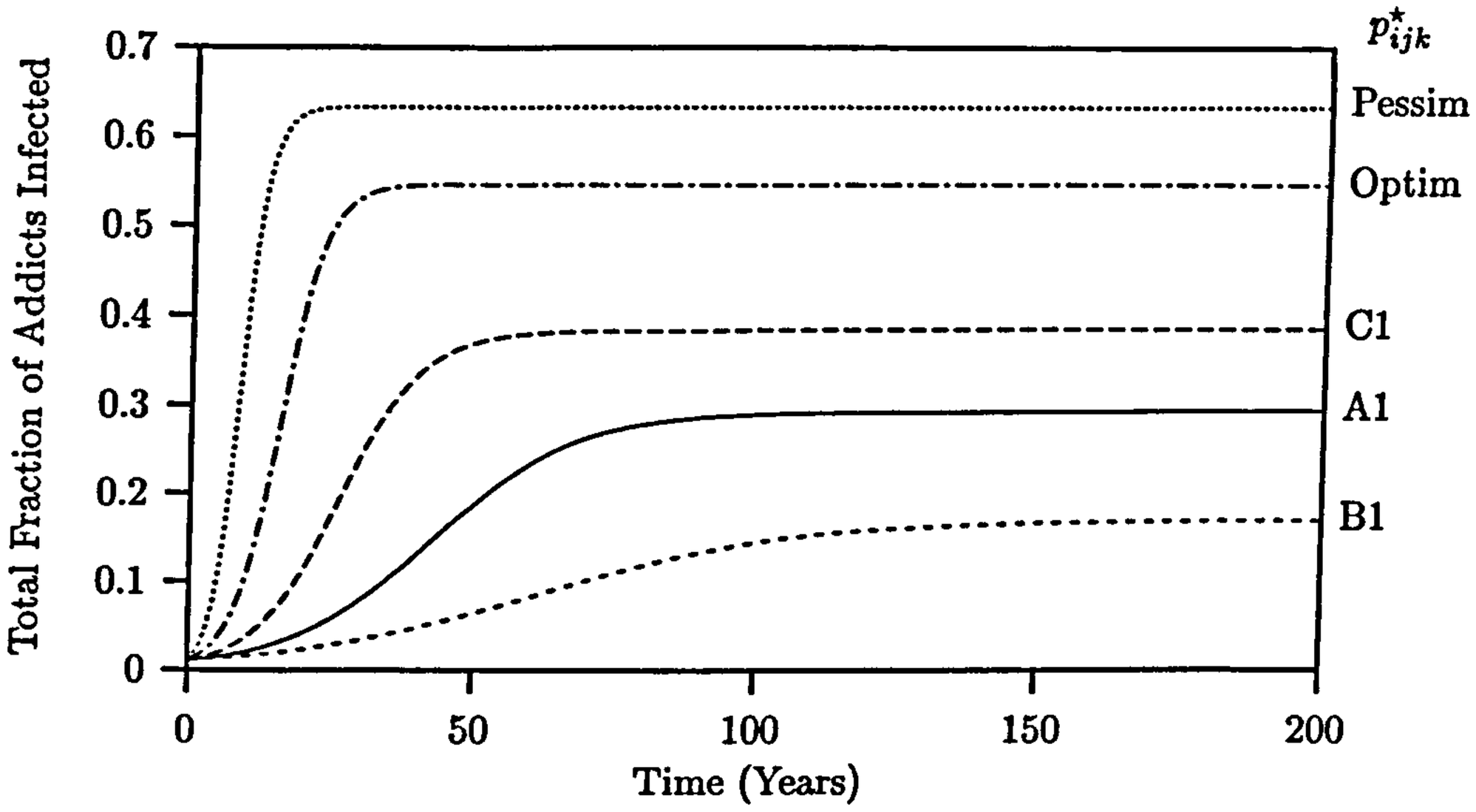


Figure 5.5: $p^{\star}_{ijk}$ Selections: Optim, A1, B1, C1, Pessim ($\pi_1(0) = 0.01$)



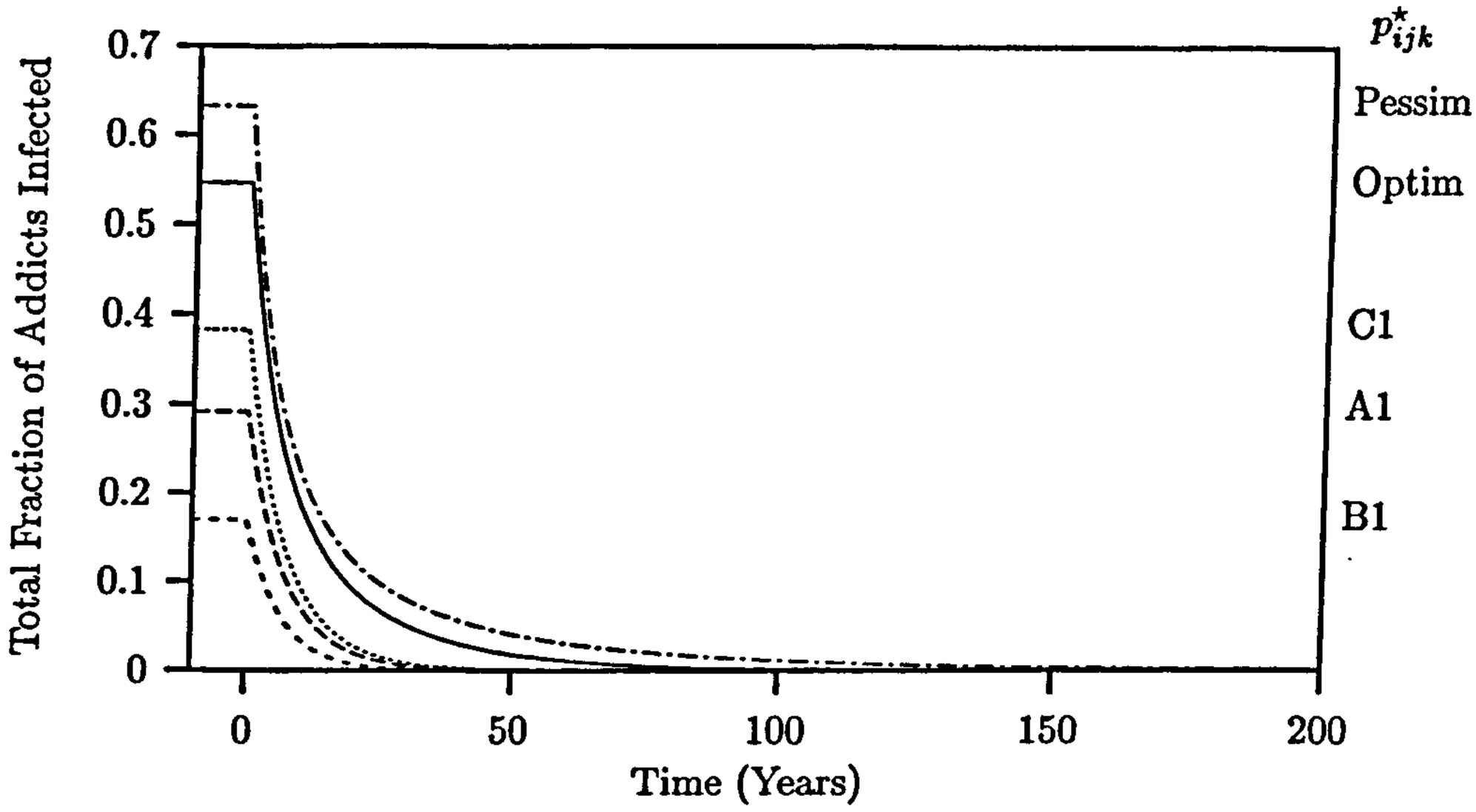
demic steady state for each set of addict-needle interactions and each of the two initial conditions. Moreover for each different set of addict-needle interaction assumptions the endemic steady state appears to be the same for both initial conditions. We have not shown simulations of the corresponding total fraction of infected needles however these are similar in behaviour to those illustrated and eventually reach an endemic steady state which is independent of the initial conditions, provided of course that disease is initially present. These figures show that as we have previously argued the total long term prevalence of disease in addicts increases as we move from the assumptions in the Optimistic Model through to the assumptions in the Pessimistic Model. If we were to illustrate any set of $p^{\star}_{ijk}$ terms (where $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$ and which satisfy the inequalities (5.1)-(5.10)) then we would expect to find that the endemic equilibrium values of these simulations lie between those of the Optimistic and Pessimistic Models.

We now illustrate simulations of the General Model using the same non-$p^{\star}_{ijk}$ parameters as in the previous simulations and the addict-needle interaction assumptions shown in Table 5.5. The $p^{\star}_{ijk}$'s in this table are the similar to those in Table 5.4 but we have now dropped the (possibly pessimistic) condition that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$. This in turn no longer requires that $p^{\star}_{211} = p^{\star}_{311} = p^{\star}_{233} = 1$. As for the previous sets of $p^{\star}_{ijk}$ terms the inequalities in (5.1)-(5.10) are again satisfied. Figure 5.5 shows simulations of the total fraction of infected addicts using the $p^{\star}_{ijk}$ terms in Table 5.5 where initially 1% of the addict population are in stage one infectivity and no other addicts or needles

Table 5.5: Addict-Needle Assumptions (No Cleaning)

$p^{\star}_{ijk}$	Optim	A1	B1	C1	Pessim
$p^{\star}_{010}$	0	0	0.25	0	0
$p^{\star}_{011}$	1	0.05	0.25	0.05	1
$p^{\star}_{012}$	0	0.3	0.25	0.3	0
$p^{\star}_{013}$	0	0.65	0.25	0.65	0
$p^{\star}_{020}$	0	0.5	0.5	0.5	0
$p^{\star}_{022}$	1	0.5	0.5	0.5	1
$p^{\star}_{030}$	0	0.1	0.33	0.1	0
$p^{\star}_{032}$	0	0.4	0.33	0.4	0
$p^{\star}_{033}$	1	0.5	0.33	0.5	1
$p^{\star}_{100}$	1	0.7	0.25	0.05	0
$p^{\star}_{101}$	0	0.05	0.25	0.7	1
$p^{\star}_{102}$	0	0.2	0.25	0.05	0
$p^{\star}_{103}$	0	0.05	0.25	0.2	0
$p^{\star}_{121}$	0	0.1	0.33	0.7	1
$p^{\star}_{122}$	1	0.7	0.33	0.1	0
$p^{\star}_{123}$	0	0.2	0.33	0.2	0
$p^{\star}_{131}$	0	0.3	0.5	0.7	1
$p^{\star}_{133}$	1	0.7	0.5	0.3	0
$p^{\star}_{200}$	1	0.7	0.5	0.3	0
$p^{\star}_{202}$	0	0.3	0.5	0.7	1
$p^{\star}_{211}$	1	0.7	0.33	0.3	1
$p^{\star}_{212}$	0	0.1	0.33	0.3	0
$p^{\star}_{213}$	0	0.2	0.33	0.4	0
$p^{\star}_{232}$	0	0.3	0.5	0.5	0
$p^{\star}_{233}$	1	0.7	0.5	0.5	1
$p^{\star}_{300}$	1	0.7	0.33	0.1	0
$p^{\star}_{302}$	0	0.2	0.33	0.2	0
$p^{\star}_{303}$	0	0.1	0.33	0.7	1
$p^{\star}_{311}$	1	0.7	0.5	0.3	1
$p^{\star}_{313}$	0	0.3	0.5	0.7	0
$p^{\star}_{322}$	1	0.7	0.5	0.3	0
$p^{\star}_{323}$	0	0.3	0.5	0.7	1

Figure 5.6: $p^{\star}_{ijk}$ Selections: Optim, A1, B1, C1, Pessim with $R_0 < 1$



are infectious. The values for R_0 in these simulations are 1.38, 1.20 and 1.70 for A1, B1 and C1 respectively. This figure clearly shows that the Optimistic Model is no longer a lower bound and moreover these simulations suggest that different interaction assumptions have a large impact on the spread of disease once the $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$ condition is lifted. Figure 5.6 uses the same five sets of addict-needle interaction assumptions as in Figure 5.5 but we now take $\phi = 0.87$ which reduces R_0 to less than unity in each simulation. For example R_0 in the Pessimistic Model is now 0.908 and therefore the values of R_0 for the other sets of p_{ijk} terms will be less than 0.908. In Figure 5.6 we have assumed that initially each model is in the endemic steady state shown in Figure 5.5 and that just after $t = 0$ years R_0 is decreased to below the critical threshold of one. It is clear from the figure that the disease eventually dies out in all addicts for each different set of interaction assumptions.

In summary, we have illustrated a small number of simulations of the Restricted General Model and the General Model. These simulations (together with many others not illustrated) suggest that when $R_0 > 1$ and disease is initially present the models tend to a stable endemic equilibrium and when $R_0 \leq 1$ disease dies out. This appears to be the case irrespective of the individual parameter estimates or initial conditions in our models.

5.8 Summary of Results for the General Mixing Model

We began this chapter by deriving the Restricted General Model, a model which can incorporate a wide range of different addict-needle interaction assumptions but which still treats interactions between infectious addicts and uninfectious needles as in the Optimistic and Pessimistic Models. We demonstrated that the Optimistic and Pessimistic Models are special cases of the Restricted General Model where the Optimistic Model represents the best case scenario and the Pessimistic Model the worst case scenario. We then stated the equations which define a model with a fully general addict-needle interaction structure, we referred to this model as the General Model. We derived the basic reproductive number for the General Model and investigated its behaviour using analytical results.

We showed that if $R_0 \leq 1$ then disease will die out in all addicts and all needles in the Restricted General Model irrespective of the initial state of the population. This result was also true for the General Model subject to several necessary inequalities between certain p^*_{ijk} terms. We also showed that there exists at least one strictly positive endemic equilibrium solution to the General Model if $R_0 > 1$ and if initially present disease will persist indefinitely. We finally investigated the behaviour of the General Model using simulation for a range of p^*_{ijk} values. These simulations suggest that as in the Optimistic and Pessimistic Models the prevalence of disease tends to an endemic steady state when $R_0 > 1$. These simulations also verified our assertion that (under certain conditions) addict-needle interaction assumptions less extreme than those in the Optimistic and Pessimistic Models give rise to long term HIV prevalence levels greater than in the Optimistic Model and less than that in the Pessimistic Model. Simulations also demonstrated that different addict-needle interaction assumptions can have a very large impact on the long term prevalence of disease.

In Chapters 2-5 we have examined a number of models which allow addicts to progress through three stages of infectivity prior to the onset of full blown AIDS. In each of these models we have assumed that the size of the addict population is constant. This is a rather unrealistic assumption as it is likely that recruitment of addicts into the needle sharing population will occur and moreover we expect that mortality from AIDS may play a significant part in reducing the size of the addict population. We assumed that the population remained at a constant size for reasons of simplicity. In the following chapter we take a brief look at extending our previous models to allow

the total population size of addicts to vary throughout the course of an HIV epidemic.

Chapter 6

Addict Recruitment and AIDS Mortality

6.1 Introduction

One of the major deficiencies in Kaplan's original model was that the total size of the injecting drug user population was assumed to be constant throughout the course of an AIDS epidemic. This assumption was primarily made for mathematical convenience rather than biological reality. Kaplan (1989) and Caulkins and Kaplan (1991) justify the assumption that addicts who leave the needle sharing population are immediately replaced by new susceptible addicts by saying that a similar assumption is common in epidemiological models where the population in question remains approximately constant over the time scale of the epidemic.

Caulkins and Kaplan argue that evidence suggests that HIV prevalence saturates quickly among high risk users (those who have a high sharing rate with strangers) compared to the long incubation period (about 9.8 years) of AIDS. This suggests that the prevalence of HIV may reach a quasi-steady state before the inevitable population changes due to mortality from AIDS. However considering the needle sharing population as a whole the argument for ignoring demographic changes due to AIDS mortality is less convincing. In any case there is little doubt that allowing for the recruitment of new susceptible drug users and mortality from AIDS to affect the size of the addict population is more realistic than assuming a constant population size.

We now investigate the effect of introducing recruitment and mortality due to AIDS into the Simple Model and the General Model. We do not need to consider the Op-

timistic and Pessimistic Models separately since they are special cases of the General Model. We first extend the differential equations which define the Simple Model to include mortality due to AIDS and allow for the influx of new susceptible drug addicts. We then investigate the behaviour of this model using analytical results. Next we move on to the General Model and in a similar fashion extend the differential equations which define this model to allow the total population size to vary. We then investigate the behaviour of this model using analytical results. We conclude with a simulation study of both models and examine the effect of different addict-needle interaction assumptions on the long term number of addicts infected.

We wish to establish whether the inclusion of recruitment and AIDS mortality affects the behaviour of our previous models. Previous studies of HIV and AIDS in drug users such as that by Greenhalgh (1996) suggest that allowing the population size to fluctuate in this manner will not effect the qualitative behaviour of our models.

6.2 Simple Model with Recruitment and AIDS Mortality

We now extend eqns (2.1)-(2.4) which define the Simple Model to include mortality due to AIDS and allow the recruitment of new susceptible drug addicts. As in Chapter 2 we have that the number of stage one addicts at time $t + \Delta t$

$$
\begin{aligned}
= \quad & \{\text{number of stage one addicts at time } t\} \\
& + \{(\text{number of uninfected addicts at time } t) \\
& \times (\text{fraction of addicts who inject in } [t, t+\Delta t) \text{ with} \\
& \text{an infectious needle which is not cleaned prior to use and} \\
& \text{where transmission of HIV occurs in a single injection})\} \\
& - \{\text{number of stage one infected addicts who progress into} \\
& \text{stage two infectivity or leave the sharing, injecting} \\
& \text{population in } [t, t+\Delta t)\}.
\end{aligned}
$$

However now we have that

$$
\begin{aligned}
n(t+\Delta t)\pi_1(t+\Delta t) \quad = \quad & n(t)\pi_1(t) + n(t)(1-\pi_1(t)-\pi_2(t)-\pi_3(t))\lambda\Delta t\beta(t)\alpha(1-\phi) \\
& -n(t)\pi_1(t)\Delta t(\mu+\delta_1) + o(\Delta t).
\end{aligned}
$$

Subtracting $n(t)\pi_1(t)$ from both sides, dividing by Δt and letting $\Delta t \to 0$ we deduce

that

$$\frac{d(n\pi_1)}{dt} = n\Big(1 - \sum_{i=1}^{3}\pi_i\Big)\lambda\beta\alpha(1-\phi) - n\pi_1(\mu+\delta_1).$$

Using the product rule we have that

$$\frac{d\pi_1}{dt} = \Big(1 - \sum_{i=1}^{3}\pi_i\Big)\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\pi_1 - \frac{\pi_1}{n}\frac{dn}{dt}.$$

Following a similar method it is straightforward to adjust eqns (2.2)-(2.3) to include a variable population size. Hence we have that

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2 - \frac{\pi_2}{n}\frac{dn}{dt},$$

and

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3 - \frac{\pi_3}{n}\frac{dn}{dt}.$$

The number of infectious needles at time $t+\Delta t$

$$\begin{aligned} &= \{\text{number of infected needles at time } t\} \\ &\quad +\{(\text{number of uninfected needles at time } t) \times (\text{fraction of} \\ &\quad \text{needles used by infected addicts in } [t, t+\Delta t))\} \\ &\quad -\{(\text{number of infected needles at time } t) \times (\text{fraction} \\ &\quad \text{of infected needles used by uninfectious addicts in } [t, t+\Delta t) \\ &\quad \text{and left in an uninfected state})\} \\ &\quad -\{\text{number of infected needles exchanged in } [t, t+\Delta t)\}. \end{aligned}$$

We now have that

$$\begin{aligned} m(t+\Delta t)\beta(t+\Delta t) = {} & m(t)\beta(t) + m(t)(1-\beta(t))\lambda\Delta t\gamma(\pi_1(t)+\pi_2(t)+\pi_3(t)) \\ & -m(t)\beta(t)\lambda\Delta t\gamma(1-\pi_1(t)-\pi_2(t)-\pi_3(t))(1-(1-\phi)(1-\theta)) \\ & -m(t)\beta(t)\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m(t)\beta(t)$ from both sides, dividing by Δt and letting $\Delta t \to 0$ we deduce that

$$\frac{d(m\beta)}{dt} = m(1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - m\beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - m\beta\tau.$$

We now assume that the total number of needles in the population at time t, $m(t)$, is proportional to the total number of addicts in the population at time t, $n(t)$. This

is a consequence of assuming a homogeneous population where all addicts behave in a similar way and so possess a similar number of needles. A similar assumption was made by Greenhalgh (1996). Letting $m(t) = n(t)/\gamma$, where γ is a constant, we have that

$$\frac{d(n\beta)}{dt} = n(1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - n\beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - n\beta\tau,$$

and using an argument similar to that in the addict equations gives us

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - \beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - \beta\tau - \frac{\beta}{n}\frac{dn}{dt}.$$

In order to define dn/dt we follow previous work by Caulkins and Kaplan (1991). If the current drug using population is of size n then new drug users enter the population at a rate proportional to n^{ν}, where ν lies strictly between zero and one. ν=0 corresponds to a constant recruitment rate as in the models of sexual transmission of HIV and AIDS among homosexuals discussed by Blythe and Anderson (1988b) and Anderson and May (1991). $\nu = 1$ corresponds to a situation where new drug users are introduced into the drug injecting population by existing drug users. When addicts have developed full blown AIDS they leave the sharing, injecting population. Only addicts in stage three infectivity can develop full blown AIDS and this occurs with rate δ_3. Addicts can also leave the sharing, injecting population for reasons other than developing AIDS, with rate μ. Hence the system of equations which describes the spread of disease among a population of varying size with three types of infectious addicts and a single type of infectious needle is:

$$\frac{d\pi_1}{dt} = \Big(1-\sum_{i=1}^{3}\pi_i\Big)\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\pi_1 - \frac{\pi_1}{n}\frac{dn}{dt}, \tag{6.1}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2 - \frac{\pi_2}{n}\frac{dn}{dt}, \tag{6.2}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3 - \frac{\pi_3}{n}\frac{dn}{dt}, \tag{6.3}$$

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - \beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - \beta\tau - \frac{\beta}{n}\frac{dn}{dt}, \tag{6.4}$$

and

$$\frac{dn}{dt} = cn^{\nu} - \mu n - \delta_3\pi_3 n, \tag{6.5}$$

with suitable initial conditions: $\pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$ and $\beta(0) \geq 0$ with $\pi_1(0) + \pi_2(0) + \pi_3(0) \leq 1$ and $\beta(0) \leq 1$. The parameter c in eqn (6.5) is chosen such that the population

size is constant in the absence of HIV, hence $c = \mu n_0^{\star 1-\nu}$ and we also assume that $0 < \nu < 1$.

Using eqns (6.1)-(6.5) we can make an important observation. We have that the equilibrium equations are identical to those for the constant population size Simple Model. We now have the addition of an equilibrium population size equation, however the values of $\pi_1^\star$, $\pi_2^\star$, $\pi_3^\star$ and $\beta^\star$ are independent of this equation. This means that we already know that there exists a unique endemic equilibrium solution if and only if $R_0 > 1$ and moreover this endemic equilibrium solution will be the same as that in the constant population size Simple Model. However since we no longer have a constant population size it is not the fraction of infected addicts and infected needles at equilibrium that we are primarily interested in but rather the *actual number* of infected addicts and needles. In addition we have that R_0 for the constant population size case will be identical to the variable population size case, this is because R_0 is only concerned with the initial exponential growth stage of an epidemic and at this early stage mortality due to the disease is irrelevant due to the small number of infectives initially present. Note also that the recruitment of susceptible addicts into the population does not affect the value of R_0.

We have that the variable population size Simple Model shares many of its characteristics with its simpler constant population size equivalent. However we do not yet know whether the stability of the equilibrium solutions in these models is the same. In other words does the disease always die out when $R_0 \leq 1$ and always tend to a unique endemic level when $R_0 > 1$ and disease is initially present? The variable population size model is much more difficult to deal with analytically than the constant population case. Our main analytical result for the variable population size Simple Model is summarised in the following theorem:

Theorem 6.1 *For the variable population size Simple Model if $R_0 \leq 1$ then there is only the disease-free equilibrium $(\pi_1, \pi_2, \pi_3, \beta, n) = (0, 0, 0, 0, n_0^\star)$ where $n_0^\star = (\frac{\mu}{c})^{\frac{1}{\nu-1}}$. If $R_0 > 1$ then as well as the disease-free equilibrium there is a unique endemic equilibrium $(\pi_1, \pi_2, \pi_3, \beta, n) = (\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star, n_e^\star)$, where $\pi_1^\star$, $\pi_2^\star$, $\pi_3^\star$ and $\beta^\star$ are the equilibrium values for the constant population size Simple Model and $n_e^\star = \left(\frac{\mu+\delta_3\pi_3^\star}{c}\right)^{\frac{1}{\nu-1}}$. The disease-free equilibrium is locally asymptotically stable if $R_0 < 1$ and unstable when $R_0 > 1$.*

Proof.

We have already shown the equilibrium results. We now show the local stability results. Substituting for $\frac{dn}{dt}$ from eqn (6.5) into eqns (6.1)-(6.4) gives us the following system of equations:

$$\frac{d\pi_1}{dt} = \Big(1-\sum_{i=1}^{3}\pi_i\Big)\lambda\beta\alpha(1-\phi) - (\mu+\delta_1)\pi_1 - \pi_1\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \quad (6.6)$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2 - \pi_2\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \quad (6.7)$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3 - \pi_3\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \quad (6.8)$$

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\Big(\sum_{i=1}^{3}\pi_i\Big) - \beta\lambda\gamma\Big(1-\sum_{i=1}^{3}\pi_i\Big)(1-(1-\theta)(1-\phi)) - \beta\tau$$

$$-\beta\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \quad (6.9)$$

and $$\frac{dn}{dt} = cn^{\nu} - \mu n - \delta_3\pi_3 n. \quad (6.10)$$

Consider the matrix **P**, the Jacobian matrix of the system represented by eqns (6.6)-(6.10) evaluated at $(0,0,0,0,n_0^\star)$,

$$\mathbf{P} = \begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda\alpha(1-\phi) & 0 \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 \\ \lambda\gamma & \lambda\gamma & \lambda\gamma & -(\lambda\gamma\hat{\theta}+\tau) & 0 \\ 0 & 0 & -n_0^\star\delta_3 & 0 & \nu c n_0^{\star(\nu-1)} - \mu \end{bmatrix}.$$

We need to show that all roots of the characteristic equation of **P** have strictly negative real parts when $R_0 < 1$. We know that the constant population model has a globally stable disease-free equilibrium if and only if $R_0 \leq 1$. Global stability implies local stability, hence all roots of the characteristic equation of

$$\mathbf{P}^\dagger = \begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda\alpha(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 \\ \lambda\gamma & \lambda\gamma & \lambda\gamma & -(\lambda\gamma\hat{\theta}+\tau) \end{bmatrix},$$

must have real parts less than or equal to zero. We require that the real parts are strictly negative. Suppose that $\omega = 0$ is a root of characteristic equation of $\mathbf{P}^\dagger$,

$$\omega^4 + a_1\omega^3 + a_2\omega^2 + a_3\omega + a_4 = 0.$$

Hence we must have that $a_4 = 0$, however from the proof of Theorem 2.3 we know that

$$a_4 = (\mu + \delta_1)(\mu + \delta_2)(\mu + \delta_3)(\lambda\gamma\hat{\theta} + \tau)(1 - R_0).$$

Hence when $R_0 < 1$, $\omega = 0$ cannot be a root and hence all roots have strictly negative real parts. If $R_0 > 1$ then we find that $a_4 = \omega_1\omega_2\omega_3\omega_4 < 0$. Either zero or two roots are complex. If all roots are real they cannot all be negative (or $a_4 \geq 0$), hence at least one root is positive. If two roots are complex then the real roots cannot both be negative (or $a_4 \geq 0$), hence at least one is positive. To find the characteristic equation of $\mathbf{P}$ we can solve

$$(\nu c n_0^{\star(\nu-1)} - \mu - \omega) \det |\mathbf{P}^\dagger - \omega \mathbf{I}| = 0. \tag{6.11}$$

Hence for $R_0 > 1$ we have that at least one root of eqn (6.11) must have a strictly positive real part. We know that for $R_0 < 1$ four of the roots of eqn (6.11) have strictly negative real parts, the remaining root being

$$\begin{aligned} \omega_5 &= \nu c n_0^{\star(\nu-1)} - \mu, \\ &= \mu(\nu - 1), \qquad \text{since } n_0^\star = \left(\frac{\mu}{c}\right)^{\frac{1}{\nu-1}}. \end{aligned}$$

Hence since $\nu < 1$ all roots have strictly negative real parts when $R_0 < 1$. This completes the proof Theorem 6.1.•

We now move on to discussing the General Mixing Model with the addition of addict recruitment and mortality due to AIDS. We examine this model in a similar fashion to the previous model.

6.3 General Model with Recruitment and AIDS Mortality

Following a similar argument to that in Section 6.2 we now extend the General Model defined by eqns (5.17)-(5.22) to allow the population size to fluctuate due to mortality from AIDS and the recruitment of susceptible drug addicts. The following system of equations describes the spread of the disease among a population of varying size with three types of infectious addicts and three types of infectious needles and general addict-needle interaction assumptions:

$$\begin{aligned} \frac{d\pi_1}{dt} &= \Big(1 - \sum_{i=1}^{3} \pi_i\Big)\lambda(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1 - \phi) - (\mu + \delta_1)\pi_1 \\ &\quad -\pi_1\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \end{aligned} \tag{6.12}$$

$$\frac{d\pi_2}{dt} = \delta_1\pi_1 - (\mu+\delta_2)\pi_2 - \pi_2\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \tag{6.13}$$

$$\frac{d\pi_3}{dt} = \delta_2\pi_2 - (\mu+\delta_3)\pi_3 - \pi_3\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big), \tag{6.14}$$

$$\begin{aligned}\frac{d\beta_1}{dt} = {} & \lambda\gamma\pi_1(1-\beta)p_{011} + \lambda\gamma\beta_2\pi_1 p_{211} + \lambda\gamma\beta_3\pi_1 p_{311} - \lambda\gamma\beta_1(1-\pi)(1-p_{101}) \\ & -\lambda\gamma\beta_1\pi_2(1-p_{121}) - \lambda\gamma\beta_1\pi_3(1-p_{131}) - \lambda\gamma\beta_1\pi_1(1-p_{111}) - \beta_1\tau \\ & -\beta_1\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big),\end{aligned} \tag{6.15}$$

$$\begin{aligned}\frac{d\beta_2}{dt} = {} & \lambda\gamma\pi_2(1-\beta)p_{022} + \lambda\gamma\beta_1\pi_2 p_{122} + \lambda\gamma\beta_3\pi_2 p_{322} + \lambda\gamma\beta_1(1-\pi)p_{102} \\ & +\lambda\gamma\beta_3(1-\pi)p_{302} + \lambda\gamma\pi_1(1-\beta)p_{012} + \lambda\gamma\pi_3(1-\beta)p_{032} + \lambda\gamma\beta_1\pi_1 p_{112} \\ & +\lambda\gamma\beta_1\pi_3 p_{132} + \lambda\gamma\beta_3\pi_1 p_{312} + \lambda\gamma\beta_3\pi_3 p_{332} - \lambda\gamma\beta_2(1-\pi)(1-p_{202}) \\ & -\lambda\gamma\beta_2\pi_1(1-p_{212}) - \lambda\gamma\beta_2\pi_2(1-p_{222}) - \lambda\gamma\beta_2\pi_3(1-p_{232}) - \beta_2\tau \\ & -\beta_2\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big),\end{aligned} \tag{6.16}$$

$$\begin{aligned}\frac{d\beta_3}{dt} = {} & \lambda\gamma\pi_3(1-\beta)p_{033} + \lambda\gamma\beta_1\pi_2 p_{123} + \lambda\gamma\beta_1\pi_3 p_{133} + \lambda\gamma\beta_2\pi_1 p_{213} + \lambda\gamma\beta_2\pi_3 p_{233} \\ & +\lambda\gamma\beta_1(1-\pi)p_{103} + \lambda\gamma\pi_1(1-\beta)p_{013} + \lambda\gamma\beta_1\pi_1 p_{113} - \lambda\gamma\beta_3(1-\pi)(1-p_{303}) \\ & -\lambda\gamma\beta_3\pi_1(1-p_{313}) - \lambda\gamma\beta_3\pi_2(1-p_{323}) - \lambda\gamma\beta_3\pi_3(1-p_{333}) - \beta_3\tau \\ & -\beta_3\Big(cn^{\nu-1} - \mu - \delta_3\pi_3\Big),\end{aligned} \tag{6.17}$$

and

$$\frac{dn}{dt} = cn^{\nu} - \mu n - \delta_3\pi_3 n, \tag{6.18}$$

again with suitable initial conditions: $0 \leq \pi_1(0)$, $\pi_2(0)$, $\pi_3(0)$, $\beta_1(0)$, $\beta_2(0)$, $\beta_3(0)$, $\pi_1(0)+\pi_2(0)+\pi_3(0) \leq 1$ and $\beta_1(0)+\beta_2(0)+\beta_3(0) \leq 1$. As in the Simple Model the parameter $c = \mu(n_0^*)^{1-\nu}$ in eqn (6.18) is such that the population size is constant in the absence of HIV.

It is clear that the equilibrium equations in the variable population size General Model are the same as those for the constant population size case. Hence as in Section 6.2 we have that the equilibrium solutions in both the constant population size and varying population size models will be the same. However again we now have the

addition of a population size equation and it is the actual number of infectious addicts and needles which is of most interest. We now turn our attention to the stability of the equilibrium solutions. As in Section 6.2 the system of equations including a variable population size is much less tractable than the constant population size case. The following theorem contains the only analytical results we have been able to obtain for the varying population size General Model. Note that both the Optimistic and Pessimistic Models are special cases of the General Model:

Theorem 6.2 *For the variable population size General Model, if* $R_0 \leq 1$ *then there is only the disease-free equilibrium* $(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3, n) = (0, 0, 0, 0, 0, 0, n_0^\star)$ *where* $n_0^\star = (\frac{\mu}{c})^{\frac{1}{\nu-1}}$. *If* $R_0 > 1$ *then as well as the disease-free equilibrium there exists an endemic equilibrium* $(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, \beta_3, n) = (\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta_1^\star, \beta_2^\star, \beta_3^\star, n_e^\star)$, *where* $\pi_1^\star$, $\pi_2^\star$, $\pi_3^\star$, $\beta_1^\star$, $\beta_2^\star$ *and* $\beta_3^\star$ *are the equilibrium values for the constant population size General Model and* $n_e^\star = \left(\frac{\mu+\delta_3\pi_3^\star}{c}\right)^{\frac{1}{\nu-1}}$. *The disease-free equilibrium is unstable if* $R_0 > 1$ *and locally asymptotically stable if* $R_0 < 1$ *and* $p_{011} \geq \max(p_{211}, p_{311})$, $p_{011} + p_{013} \geq p_{211} + p_{213}$ *and* $p_{033} \geq p_{233}$.

Proof.

Consider the matrix **W**, the Jacobian matrix of eqns (6.12)-(6.18) evaluated at $(0, 0, 0, 0, 0, 0, n_0^\star)$, $\mathbf{W} =$

$$\begin{bmatrix}
-(\mu+\delta_1) & 0 & 0 & \lambda\alpha_1(1-\phi) & \lambda\alpha_2(1-\phi) & \lambda\alpha_3(1-\phi) & 0 \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 & 0 \\
\lambda\gamma p_{011} & 0 & 0 & -\lambda\gamma(1-p_{101}+\hat{\tau}) & 0 & 0 & 0 \\
\lambda\gamma p_{012} & \lambda\gamma p_{022} & \lambda\gamma p_{032} & \lambda\gamma p_{102} & -\lambda\gamma(1-p_{202}+\hat{\tau}) & \lambda\gamma p_{302} & 0 \\
\lambda\gamma p_{013} & 0 & \lambda\gamma p_{033} & \lambda\gamma p_{103} & 0 & -\lambda\gamma(1-p_{303}+\hat{\tau}) & 0 \\
0 & 0 & -n_0^\star\delta & 0 & 0 & 0 & \nu c n_0^{\star(\nu-1)} - \mu
\end{bmatrix}.$$

We know that the constant population size General Model has a globally stable disease-free equilibrium if $R_0 \leq 1$ and $p_{011} \geq \max(p_{211}, p_{311})$, $p_{011} + p_{013} \geq p_{211} + p_{213}$ and $p_{033} \geq p_{233}$. Hence all roots of the characteristic equation of $\mathbf{W}^\dagger =$

$$\begin{bmatrix} -(\mu+\delta_1) & 0 & 0 & \lambda\alpha_1(1-\phi) & \lambda\alpha_2(1-\phi) & \lambda\alpha_3(1-\phi) \\ \delta_1 & -(\mu+\delta_2) & 0 & 0 & 0 & 0 \\ 0 & \delta_2 & -(\mu+\delta_3) & 0 & 0 & 0 \\ \lambda\gamma p_{011} & 0 & 0 & -\lambda\gamma(1-p_{101}+\hat{\tau}) & 0 & 0 \\ \lambda\gamma p_{012} & \lambda\gamma p_{022} & \lambda\gamma p_{032} & \lambda\gamma p_{102} & -\lambda\gamma(1-p_{202}+\hat{\tau}) & \lambda\gamma p_{302} \\ \lambda\gamma p_{013} & 0 & \lambda\gamma p_{033} & \lambda\gamma p_{103} & 0 & -\lambda\gamma(1-p_{303}+\hat{\tau}) \end{bmatrix},$$

must have real parts less than or equal to zero. We require that the real parts are strictly negative. The characteristic equation of $\mathbf{W}^\dagger$ is

$$\omega^6 + a_1\omega^5 + a_2\omega^4 + a_3\omega^3 + a_4\omega^2 + a_5\omega + a_6 = 0.$$

From the proof of Lemma 5.4 we know that

$$a_6 = (\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)\lambda^3\gamma^3(1-p_{101}+\hat{\tau})(1-p_{202}+\hat{\tau})(1-p_{303}+\hat{\tau})(1-R_0).$$

Hence when $R_0 < 1$, $\omega = 0$ cannot be a root and hence all roots have strictly negative real parts. If $R_0 > 1$ then we find that $a_6 = \omega_1\omega_2\omega_3\omega_4\omega_5\omega_6 < 0$ which implies that at least one root must have a strictly positive real part by a similar argument to the corresponding result for the variable population size Simple Model. To find the characteristic equation of $\mathbf{W}$ we solve

$$(\nu c n_0^{\star(\nu-1)} - \mu - \omega)\det|\mathbf{W}^\dagger - \omega\mathbf{I}| = 0. \tag{6.19}$$

Hence for $R_0 > 1$ we have that at least one root of eqn (6.19) must have a strictly positive real part. We know that for $R_0 < 1$ and $p_{011} \geq \max(p_{211}, p_{311})$, $p_{011} + p_{013} \geq p_{211} + p_{213}$ and $p_{033} \geq p_{233}$, six of the roots of eqn (6.19) have strictly negative real parts, the remaining root being $\omega_7 = \nu c n_0^{\star(\nu-1)} - \mu = \mu(\nu - 1)$. Hence since $\nu < 1$ all roots have strictly negative real parts when $R_0 < 1$. This completes the proof of Theorem 6.2.•

Theorems 6.1 and 6.2 suggest that as in the constant population size models, $R_0 = 1$ is the critical threshold between the disease taking off or dying out. However these analytical results are limited to local behaviour about the disease-free equilibrium. To examine the global behaviour of our models we need to use numerical simulation.

6.4 Simulation Study

We now conduct a brief simulation study of our two variable population size models. We expect that introducing recruitment and AIDS mortality will not effect the qualitative behaviour of our previous (constant population size) models and they will still reach an endemic equilibrium when $R_0 > 1$ (and disease is initially present) and the disease-free equilibrium when $R_0 \leq 1$. We do however expect that the time taken for the disease to reach equilibrium will be increased due to the additional demographic processes which have been introduced. This behaviour is consistent with that found by Greenhalgh (1996) for heterogeneous needle sharing models.

Before we can simulate our variable population size models there are two additional parameter estimates that we require. Firstly we require a reasonable estimate of the size of a population of needle sharing drug users prior to the introduction of HIV. We follow Kaplan and O'Keefe (1993) who estimate that the size of the needle sharing population in New Haven, Connecticut, USA is approximately $n = 2,300$. It is reasonable to use the size of this population in our models as many of our other parameters are also based on the behaviour of this population. The only remaining parameter to be estimated is ν, (recall that new drug users are recruited into the population at rate cn^{ν}). As mentioned above $\nu = 1$ corresponds to each drug user recruiting his or her acquaintances into drug use whereas $\nu = 0$ corresponds to a fixed proportion of the population having an innate propensity for a given level of drug use. We take $\nu = 0.5$ as a compromise between these extreme situations. As mentioned previously the parameter $c = \mu n^{1-\nu}$ is uniquely determined by the values of μ, n and ν, hence $c = 6.3928$ per year.

Simple Model Simulations

Figure 6.1 is similar to Figure 2.1 in Chapter 2 and uses the same initial conditions ($\pi_1(0) = 0$ with all other types of addicts and needles initially zero) but now assumes that the initial size of the population is 2,300 drug users and as mentioned above $\nu = 0.5$. The figure shows the progress of the various stages of HIV over time together with the total fraction of needles contaminated with HIV. The most striking difference between the behaviour of these simulations and those in Figure 2.1 is the time taken for the prevalence of disease in each infectious stage to reach a steady state. It takes about 50 or so years for the simulations in Figure 6.1 to reach a steady state compared with only about 25 years for similar simulations of the constant population size model.

Another interesting contrasting feature of the simulations in Figure 6.1 is the way the prevalence in each infectious stage reaches a maximum value and then decreases to a steady state compared to the monotonic nature of the simulations in Figure 2.1. This is because the epidemiological processes happen on a much faster timescale than the demographic ones. The fractions of the different types of addicts and the different types of needles settle down a lot faster than the total number of addicts in the population. The approximate steady state values are $(\pi_1^\star, \pi_2^\star, \pi_3^\star, \beta^\star) = (0.027, 0.355, 0.249, 0.675)$. These are of course the same as for the constant population size model using the same parameter estimates. Figure 6.2 shows the behaviour of the total size of the population corresponding to the spread of disease shown by the simulations in Figure 6.1. After approximately seventy years the initial population size of 2,300 has been greatly reduced to a steady state value of approximately 1,044 drug users.

Figures 6.3-6.5 show the behaviour of the total fraction of infected needles and the total fraction of infected addicts for the same six initial conditions used in Figures 2.2-2.4 and the same parameter estimates. These figures clearly show that the prevalence of disease in addicts and needles eventually reaches a steady state for each set of initial conditions and moreover this steady state is the same as that in Figure 6.1 of $(\pi^\star, \beta^\star) = (0.633, 0.675)$. We have not shown the corresponding behaviour of the total population size, however for each of the different initial conditions the total population size settles down the same steady state size of 1,044 drug users shown in Figure 6.2. These simulations suggest that if disease is initially present then the prevalence of disease will tend to an endemic equilibrium when $R_0 > 1$. Simulations for a variety of different initial conditions and parameter estimates suggest that this is always the case. We have not illustrated any simulations where $R_0 \leq 1$, however simulations suggest that in this case the disease always dies out irrespective of the initial conditions or individual parameter estimates and the population size tends to the unique disease-free equilibrium.

6.4.1 General Model Simulations

Having examined the effects of including a variable population size into the Simple Model we now look at the effect of this modification in the General Model. As in Chapter 5 we are mainly interested in how the General Model behaves for various different choices of addict-needle interaction assumptions. We use the non-p_{ijk} parameters

Figure 6.1: Variable Population Size Simple Model $R_0 > 1$

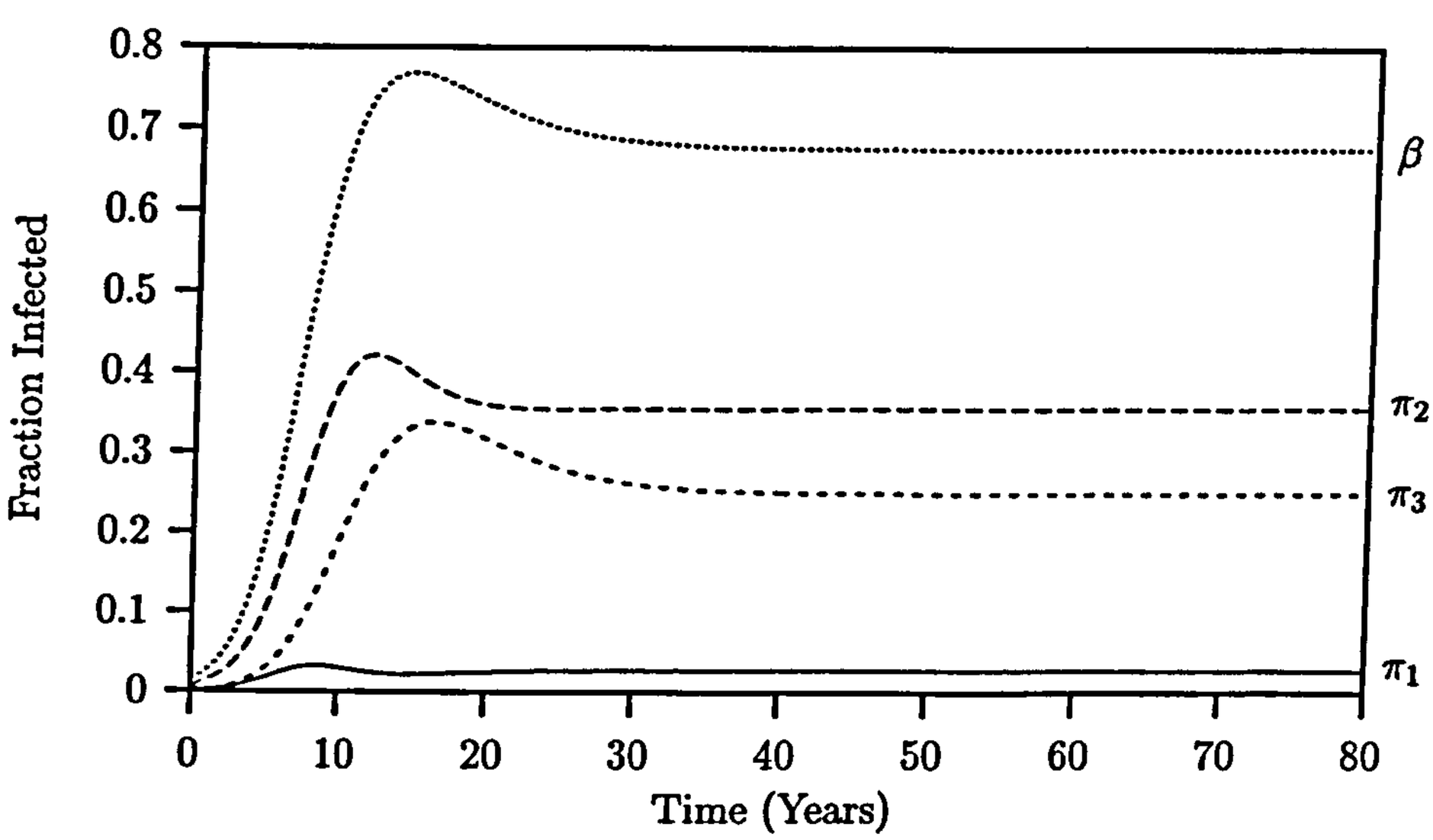


Figure 6.2: Variable Population Size Simple Model $R_0 > 1$

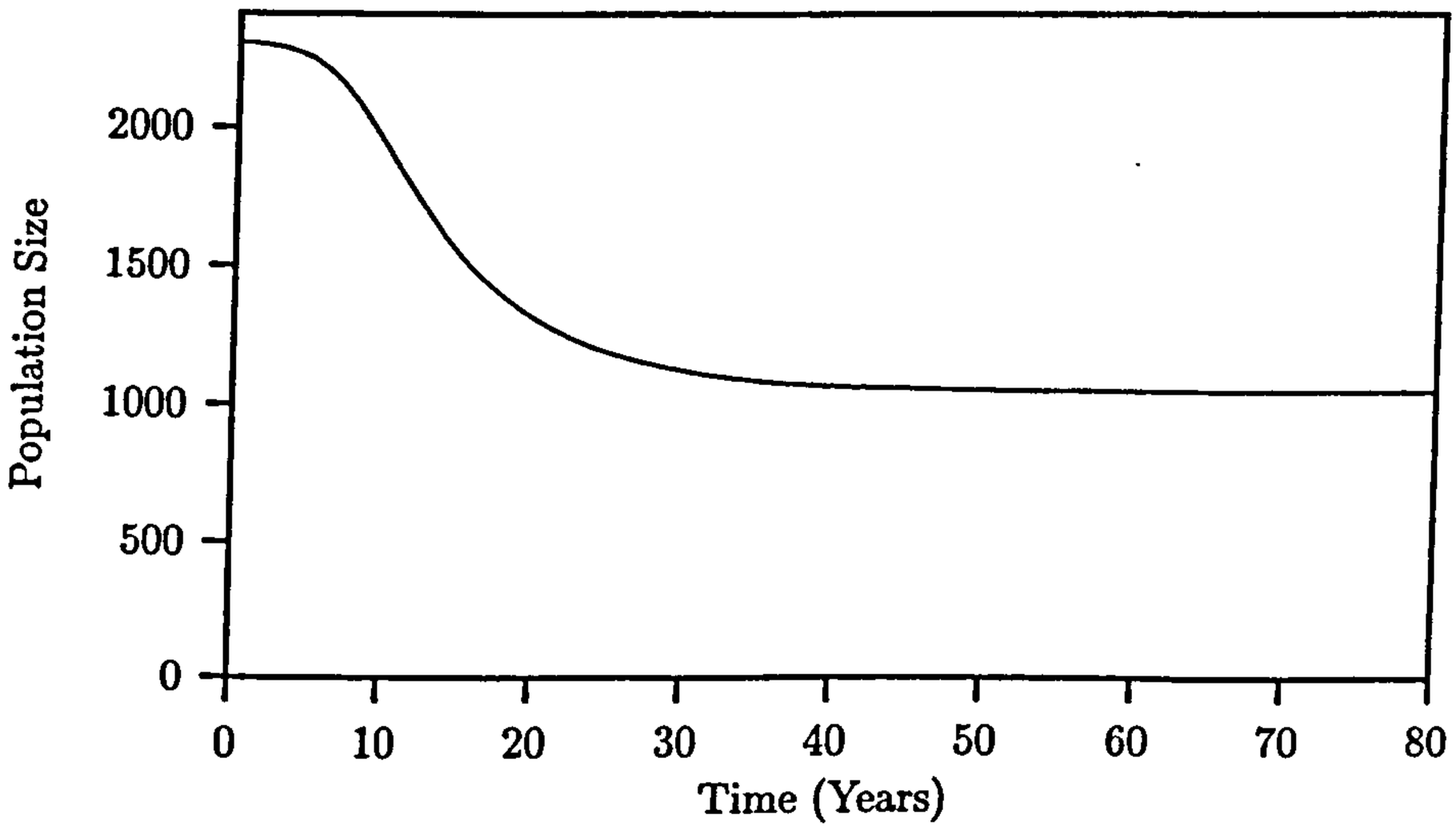


Figure 6.3: Variable Population Size Simple Model when $R_0 > 1$

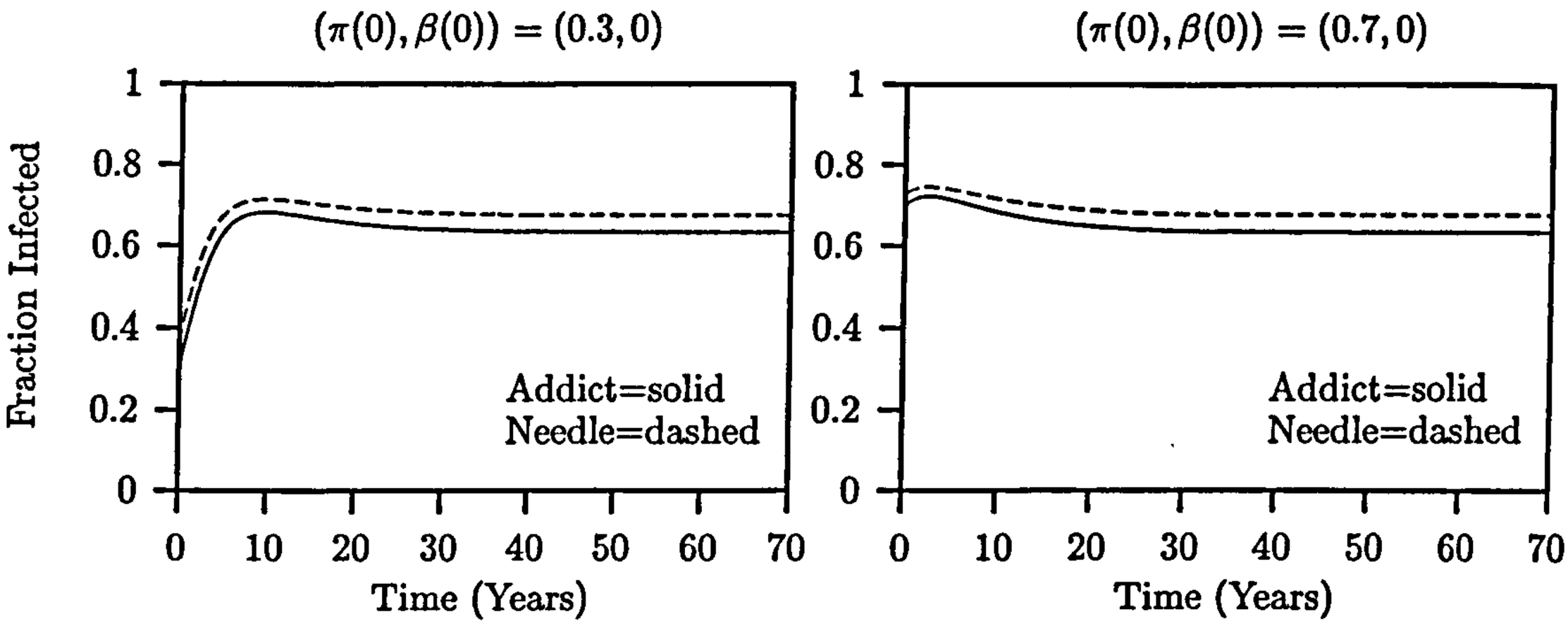


Figure 6.4: Variable Population Size Simple Model when $R_0 > 1$

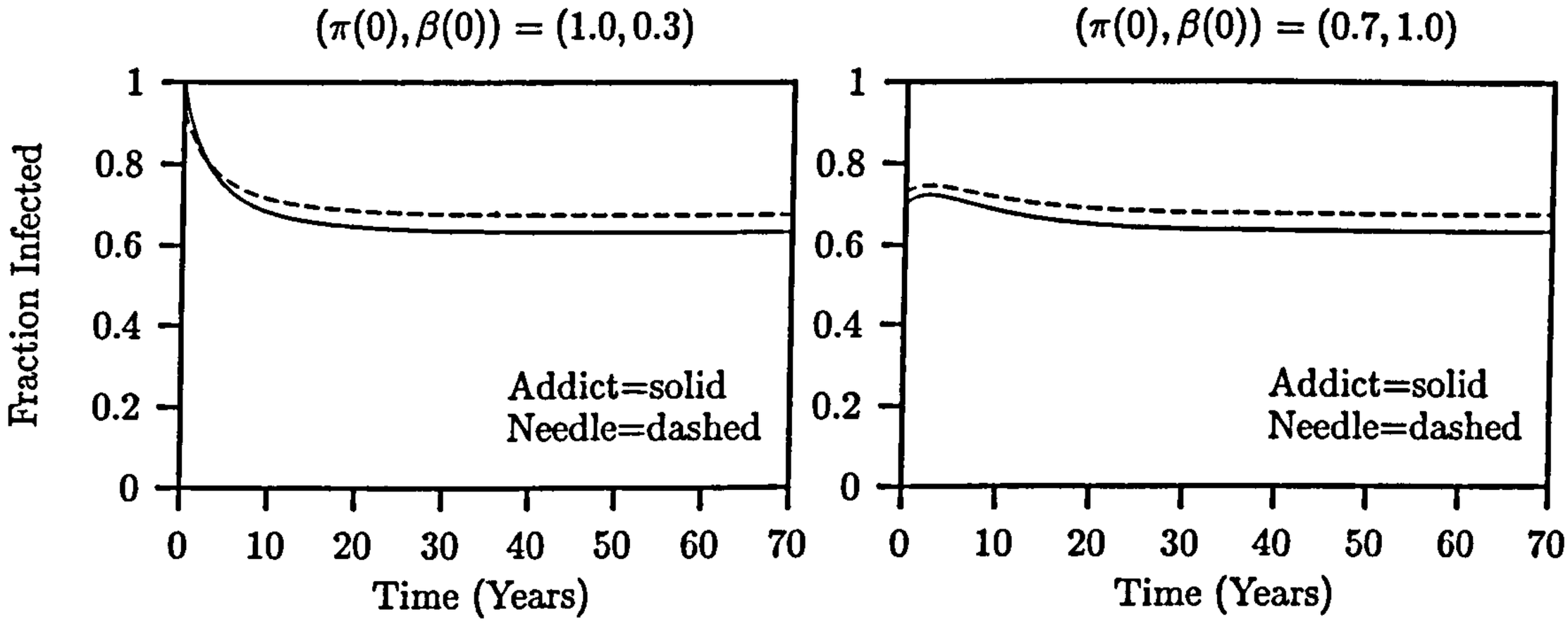
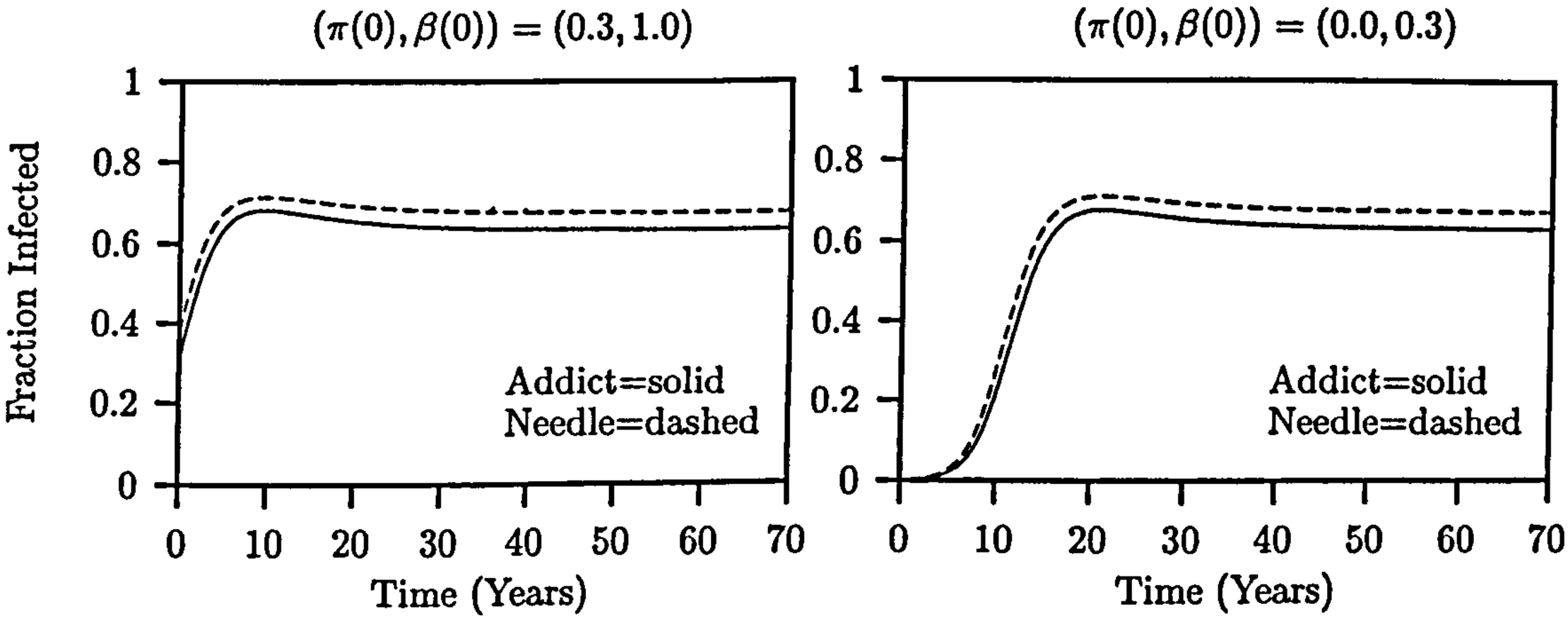


Figure 6.5: Variable Population Size Simple Model when $R_0 > 1$



outlined in Table 3.1 together with $\nu = 0.5$ and assume that initially the population consists of 2,300 addicts. This leaves us only requiring to choose the addict-needle interaction assumptions which we are interested in. To this end we use the same $p^{\star}_{ijk}$ terms as outlined in Table 5.4 in Chapter 5. Recall that this table contained five addict-needle interaction assumptions labelled Optim, A, B, C and Pessim respectively. The first and last of these contain the assumptions used in the Optimistic and Pessimistic Models respectively, those labelled A-C contain less extreme assumptions.

Figure 6.6 shows simulations of the total fraction of infected addicts for each of the five addict-needle interaction assumptions. It was initially assumed that 1% of the addict population was in stage one infectivity and no other addicts or needles were infected. The figure shows that as we might expect the Pessimistic Model gives rise to the highest long term prevalence of disease and the Optimistic Model the lowest long term prevalence of disease. As in the variable population size Simple Model these simulations take longer to reach equilibrium than those of the Restricted General Model in the previous chapter. We also have that now the prevalence of disease peaks slightly before decreasing into an equilibrium state. Figure 6.7 shows the corresponding behaviour of the total population size of addicts as the epidemic spreads among the population. Again as seems intuitive the final equilibrium population size is lowest in the Pessimistic Model and highest in Optimistic Model with the other addict-needle interaction assumptions giving values in between these two extremes.

Simulations for a variety of initial conditions and parameter estimates suggest that the General Model (both restricted and unrestricted) with addict recruitment and AIDS mortality has the same qualitative behaviour as the simpler constant population size case. We have not shown simulations where $R_0 \leq 1$ however it is again the case that disease dies out in all addicts and all needles.

6.4.2 The Number of Addicts Infected

We have previously illustrated simulations of the prevalence of disease and the corresponding effect on the total population size. We now examine simulations of the actual number of addicts infected (which is the fraction of addicts infected multiplied by the population size). When the population size is influenced by the spread of disease it is misleading to compare the prevalence of disease (for example comparing the long term prevalence between the Optimistic and Pessimistic Models) in isolation without

Figure 6.6: Total Prevalence of Disease Among Addicts

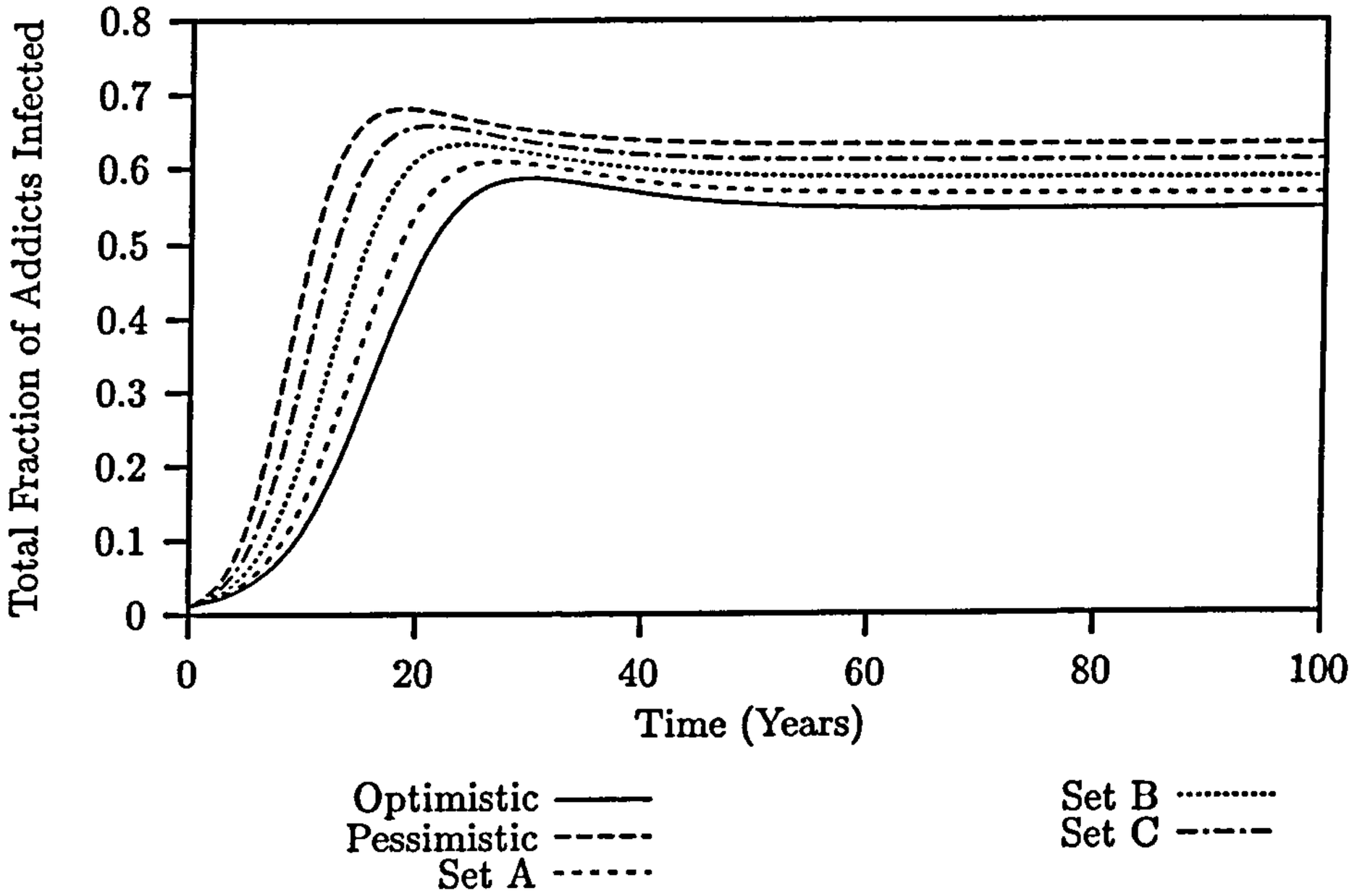


Figure 6.7: Total Population Size

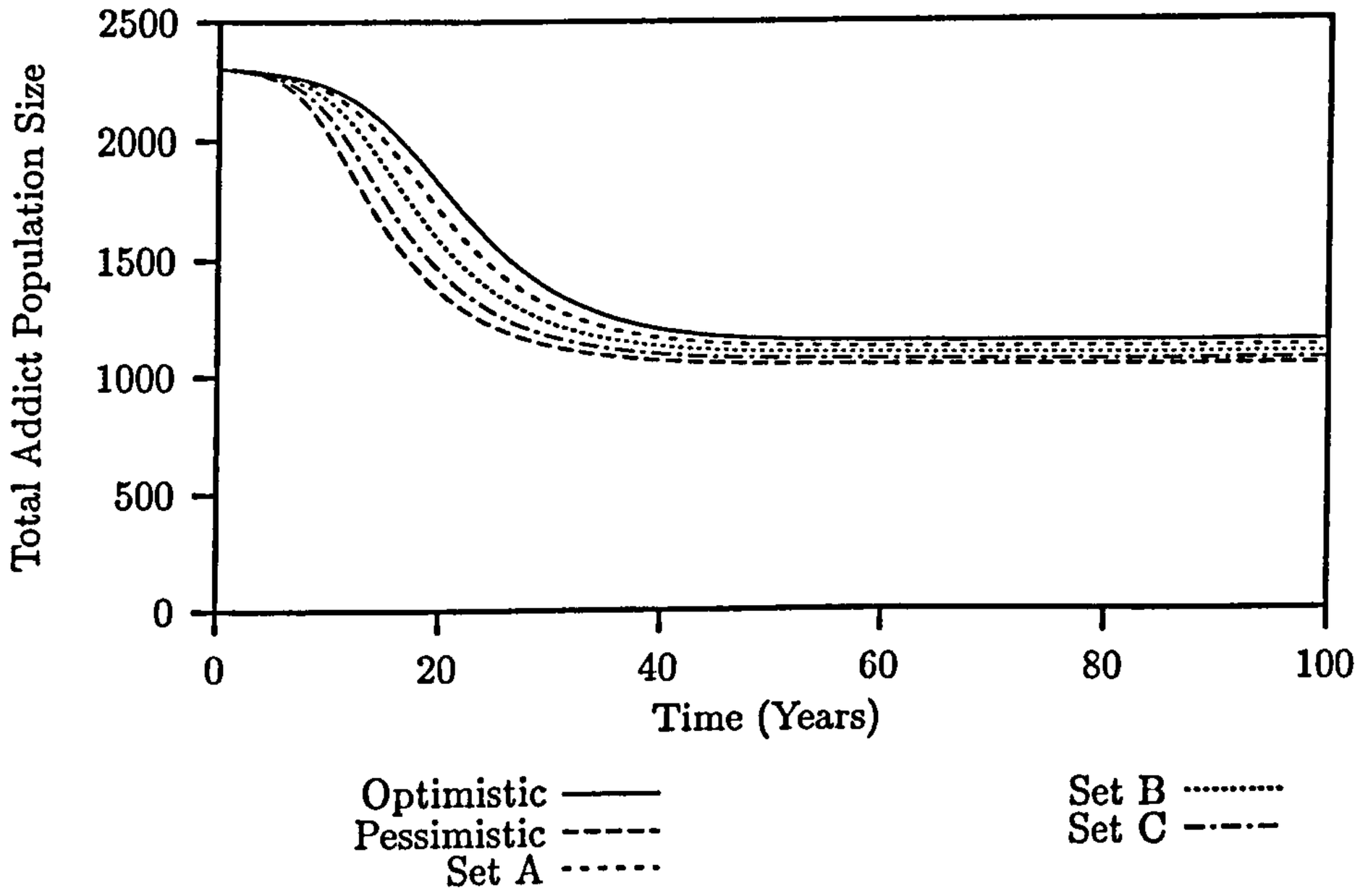
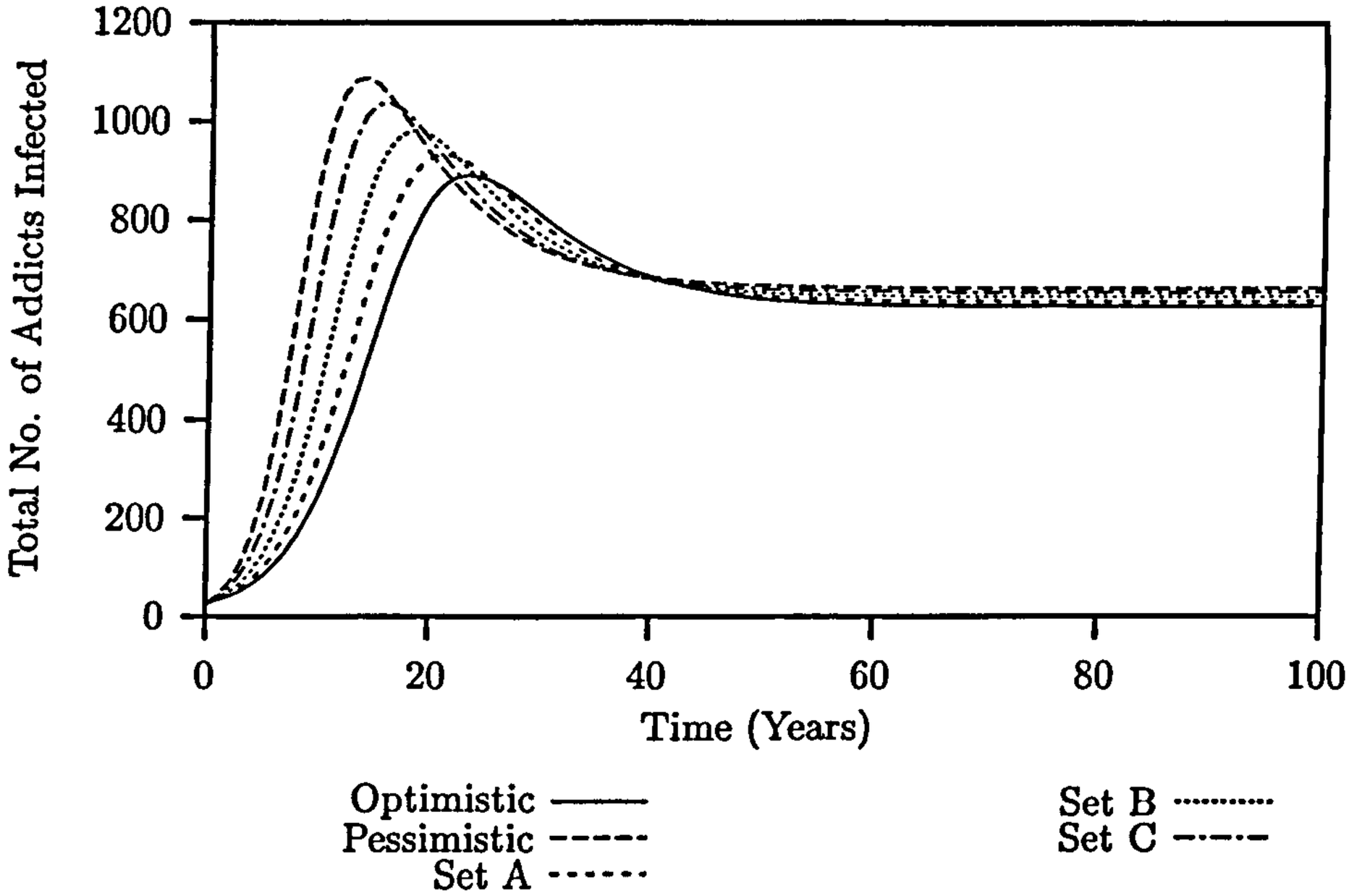


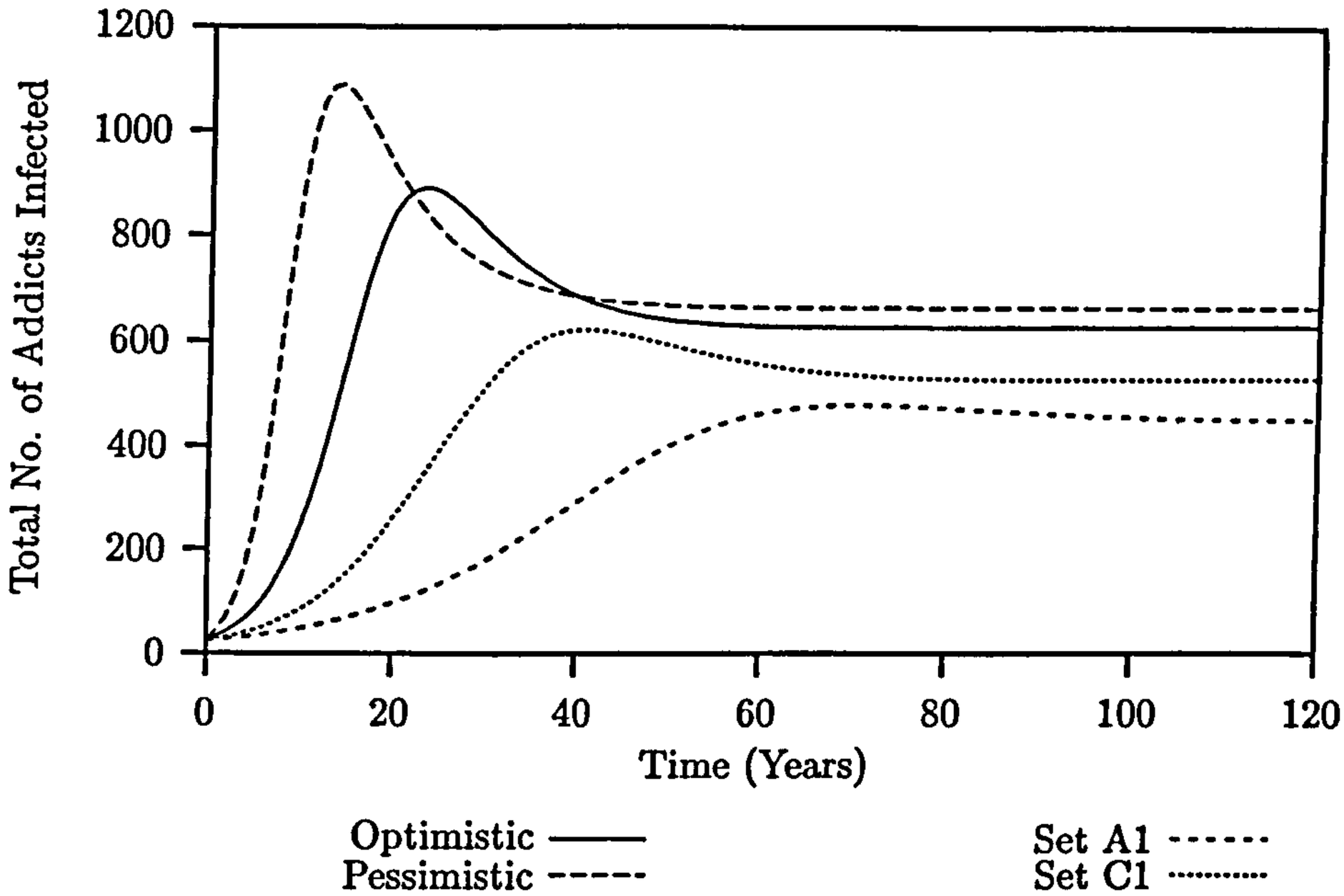
Figure 6.8: Total Number of Addicts Infected (Restricted General Model)



examining the corresponding population sizes. For example in Figure 6.7 the long term prevalence in addicts varies from about 54% to 63%, however these figures relate to populations of different sizes. The Optimistic Model gives rise to a long term prevalence of 54% among a population of 1,145 addicts whereas the Pessimistic Model gives a higher long term prevalence of 63% but this is among a smaller population size of 1,044 addicts. Figure 6.8 shows the total number of infected addicts over time in the Optimistic and Pessimistic Models and the Restricted General Model using interaction assumptions A-C. It is clear from this figure that the long term number of infected addicts is very close in each of these models, indeed much closer than suggested by the simulations in Figure 6.6. Intuitively this is because the greater the spread of disease the higher the AIDS mortality and therefore the lower the final population size. This is an important distinction between our previous constant size models and those with recruitment and AIDS mortality. For example the simulations in Figure 6.8 suggest that for the current choice of parameter estimates there is little to choose between any of the different addict-needle interaction assumptions with respect to their effect on the long term total number of addicts infected. In this case we may as well opt for the Optimistic Model as our best three stage infectivity model since this is by far the simplest model.

Figure 6.9 is similar to Figure 6.8 but illustrates the Optimistic and Pessimistic

Figure 6.9: Total Number of Addicts Infected (General Model)



Models and the General Model using addict-needle interactions A1 and C1 shown in Table 5.5. These simulations show that as suggested in Figure 5.5 in Chapter 5, once the condition $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$ is dropped then different addict-needle interaction assumptions can have a large impact on the spread of disease. Simulations (not illustrated) show that the long term prevalence of disease in addicts in each of these variable population size models is the same as illustrated in Figure 5.5. However Figure 6.9 shows that as discussed above, the actual number of addicts infected in these various models are not as different as we might have expected considering the large differences in the long term prevalence of disease between these models.

6.5 Summary of Addict Recruitment and AIDS Mortality

In this chapter we examined the effect of including mortality from AIDS and recruitment of new susceptible drug users into the Simple and General Models from Chapters 2 and 5 respectively. We found that including these features does not affect the qualitative behaviour of these models. In particular $R_0 = 1$ is still the critical threshold between the disease taking off or dying out. These findings are consistent with other studies of variable population size models. We have noted two main effects of allowing the population size to fluctuate according to the spread of disease. Firstly it appears that

the time taken for our models to reach an endemic equilibrium has increased. Secondly and more importantly our simulations suggest that the effect of different addict-needle interaction assumptions is substantially less than in our previous models where the population size remained constant.

Having examined a number of different three stage infectivity models and established various properties of these models we now pull together the results of Chapters 2-6. In the following chapter we investigate the differences between our three stage infectivity models and conventional single stage infectivity models, in particular we are interested in whether three stage infectivity gives rise to a higher (or lower) long term prevalence of disease than in single stage models.

Chapter 7

Practical Implications of Models

7.1 Introduction

In Chapters 2-6 we investigated the properties of various mathematical models of the spread of HIV via needle sharing. These models started off simply and became more complex as we relaxed particular assumptions in an effort to make our models more realistic. Our first model extended a model by Kaplan and O'Keefe to allow addicts infected with HIV to progress through different stages of infectivity. Our next two models additionally allowed needles infected with HIV to exist in different infectious states and assumed that the various types of infectious addicts and infectious needles interacted with each other in very specific ways. We then looked at the Restricted General Model and the General Model. These models also included three types of infectious addicts and three types of infectious needles but additionally allowed addicts and needles to interact with each other in a wide variety of ways. Finally we relaxed the assumption that the size of the addict population is always constant and included mortality due to AIDS and the recruitment of new susceptible drug users into our models.

In this chapter we focus on what new information can be gained from analysing the models in Chapters 2-6. It is important to remember that while some of our previous work has been rather abstract, the main purpose in constructing these models is to provide information to assist in the planning and management of the spread of HIV and AIDS among drug using populations. There are two main areas we wish to address. Firstly, we wish to determine whether models which include a three stage infectious period give rise to different long term prevalence levels of HIV than existing single stage

infectivity models. This is important since many existing HIV and AIDS models in the literature incorporate (at least in part) the spread of HIV via needle sharing. Secondly we are interested in whether incorporating a needle exchange program (as advocated by Kaplan and O'Keefe) into our three stage models will produce a decrease in the long term prevalence of HIV of a similar magnitude to that of single stage models. This is important as much of the argument for introducing needle exchange programs in the United States is based upon single stage infectivity models. This chapter aims to provide sensible and structured answers to these two important areas.

We begin with a discussion on how our three stage infectivity models should be calibrated so that they provide meaningful comparisons with existing single stage infectivity models. We derive various calibration equations and comment on which are the most appropriate for our purposes. We then discuss the effect of flushing in single stage models and addict-needle interaction assumptions in our three stage models and the effect this has on our calibration method. We then compare the single stage Kaplan and O'Keefe Model with the Restricted General Model and then finally the General Model and comment on the resulting differences between these models. Next we move on to our second area of interest and examine the effect of control strategies such as needle exchange and improved needle cleaning in our three stage models. The chapter concludes with a summary of the main points.

7.2 Calibration Method

The model by Kaplan and O'Keefe was one of the first to deal with the spread of HIV via needle sharing and the first model to examine the benefits of needle exchange programs. This model is generally accepted as a reasonable account of how HIV spreads among addicts and needles in a population of intravenous drug users. We wish to examine the difference in the long term prevalence of HIV between Kaplan and O'Keefe's original model and our three stage infectivity models. We require a calibration method to ensure that any differences in our models are caused only by the effects of allowing addicts and needles to exist in three different infectious classes. There are a number of different potential calibrations we could use. We have shown that R_0 is an important quantity in all our models and hence ensuring that R_0 is the same in the models we wish to compare might be a suitable form of calibration. By equating R_0 in our models we could then examine any differences in the steady state solutions. Alternatively another

calibration would be to ensure that the steady state solutions were the same in the models being compared which would allow us to compare R_0 values. In this section we instead calibrate our models directly on biological grounds, however as we shall see later this is equivalent to using R_0 as a calibration.

We now outline several biological differences between Kaplan and O'Keefe's model and our three stage models. Firstly for example in Kaplan and O'Keefe's model $1/\delta$ is the average AIDS incubation period, conditional on not leaving the sharing, injecting population during that period. For the Simple Model the corresponding quantity is

$$\frac{1}{\delta_1} + \frac{1}{\delta_2} + \frac{1}{\delta_3}, \tag{7.1}$$

which suggests the following calibration between these two models:

$$\frac{1}{\delta} = \frac{1}{\delta_1} + \frac{1}{\delta_2} + \frac{1}{\delta_3}. \tag{7.2}$$

However this is not the only possible calibration, nor even obviously the correct one. If leaving the sharing, injecting population for reasons other than developing full blown AIDS is taken into consideration then for Kaplan and O'Keefe's model the average time that an infected addict spends in the population is $1/(\mu + \delta)$. Using the parameter estimates in Table 2.1 (together with $\delta = 0.101958$ from Kaplan and O'Keefe, 1993) this means that addicts spend on average 4.25 years from becoming infected to leaving the population (whether due to developing full blown AIDS or other reasons). For the Simple Model the formula for the corresponding period for an infected addict is

$$\frac{1}{\mu+\delta_1} + \frac{\delta_1}{(\mu+\delta_1)(\mu+\delta_2)} + \frac{\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}. \tag{7.3}$$

Using the parameter estimates in Table 3.1 we find that addicts in the Simple Model spend on average 4.79 years from becoming infected to leaving the population (due to developing full blown AIDS or other reasons). This suggests that an alternative calibration is to make these two periods equal, namely

$$\frac{1}{\mu+\delta} = \frac{1}{\mu+\delta_1} + \frac{\delta_1}{(\mu+\delta_1)(\mu+\delta_2)} + \frac{\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}. \tag{7.4}$$

For the spread of HIV in a heterosexual population, μ, the rate at which an individual leaves the sexually mixing population for reasons other than developing full blown AIDS, might reasonably be assumed to be quite small as non-HIV related death rates are small for the age-classes where individuals are sexually mixing, so the two calibrations in eqns (7.2) and (7.4) will be approximately the same. For HIV/AIDS amongst

injecting drug users, μ is larger and as we have seen the two calibrations are not the same. It is not unreasonable to require that for a sensible comparison between the Simple Model and the Kaplan and O'Keefe Model, the average time between an addict becoming infected and leaving the sharing, injecting population (whether due to developing full blown AIDS or other reasons) should be the same which suggests that eqn (7.4) is the more appropriate calibration.

To summarise we have outlined two biological differences between Kaplan and O'Keefe's original model and our three stage models. However we do not calibrate our models using either eqn (7.2) or eqn (7.4) as there is another more relevant biological difference between these models which we now discuss.

We have constructed the Simple, Optimistic, Pessimistic and General Models to investigate the effect of allowing addicts to have a variable viral load over the course of their infectious lifetime. Therefore a natural calibration method is to ensure that the cumulative viral load over the infectious lifetime of an addict is the same in all models. We have assumed that the viral load in an infectious addict is the same as that in an infectious needle. Hence it is reasonable to assume that the HIV transmission probability α (or α_i for the models with three types of infectious needles) should be proportional to the viral load of an infectious addict. From eqn (7.4) we know how long on average an addict spends in the sharing, injecting population in each model. Hence to ensure that the cumulative viral load over the infectious lifetime of a typical addict is the same in the Kaplan and O'Keefe Model and the Simple Model we require that

$$\frac{\alpha}{\mu+\delta} = \left[\frac{1}{\mu+\delta_1} + \frac{\delta_1}{(\mu+\delta_1)(\mu+\delta_2)} + \frac{\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\right]\alpha', \qquad (7.5)$$

where α' denotes the probability of transmission of HIV from injecting once with an infectious needle in the Simple Model. In order that eqn (7.5) is satisfied we need to adjust at least one of the model parameters. Since our three stage models are concerned with allowing addicts to progress through different viral load stages it seems natural to adjust the value of α', which represents the viral load of an addict in the Simple Model. This means that we can also maintain the relationship $(1/\delta) = (1/\delta_1)+(1/\delta_2)+(1/\delta_3)$. Intuitively this relationship should always hold since we are simply splitting up the total incubation period into three shorter periods. Therefore to calibrate the Kaplan and O'Keefe Model with the Simple Model we fix all our model parameters at their estimated values with the exception of α' which we choose such that eqn (7.5) is satisfied. In a similar manner we have that the calibration equation for comparing

the Simple Model with the General Model (including the Optimistic and Pessimistic Models) is

$$\left[\frac{1}{\mu+\delta_1}+\frac{\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}+\frac{\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\right]\alpha'$$

$$= \left[\frac{\zeta_1}{\mu+\delta_1}+\frac{\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}+\frac{\zeta_3\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\right]\alpha_2, \qquad (7.6)$$

where $\alpha_1 = \zeta_1\alpha_2$, and $\alpha_3 = \zeta_3\alpha_2$. In this comparison we again fix all model parameters at their estimated values (including ζ_1 and ζ_3) with the exception of α_2 and adjust its value to satisfy eqn (7.6).

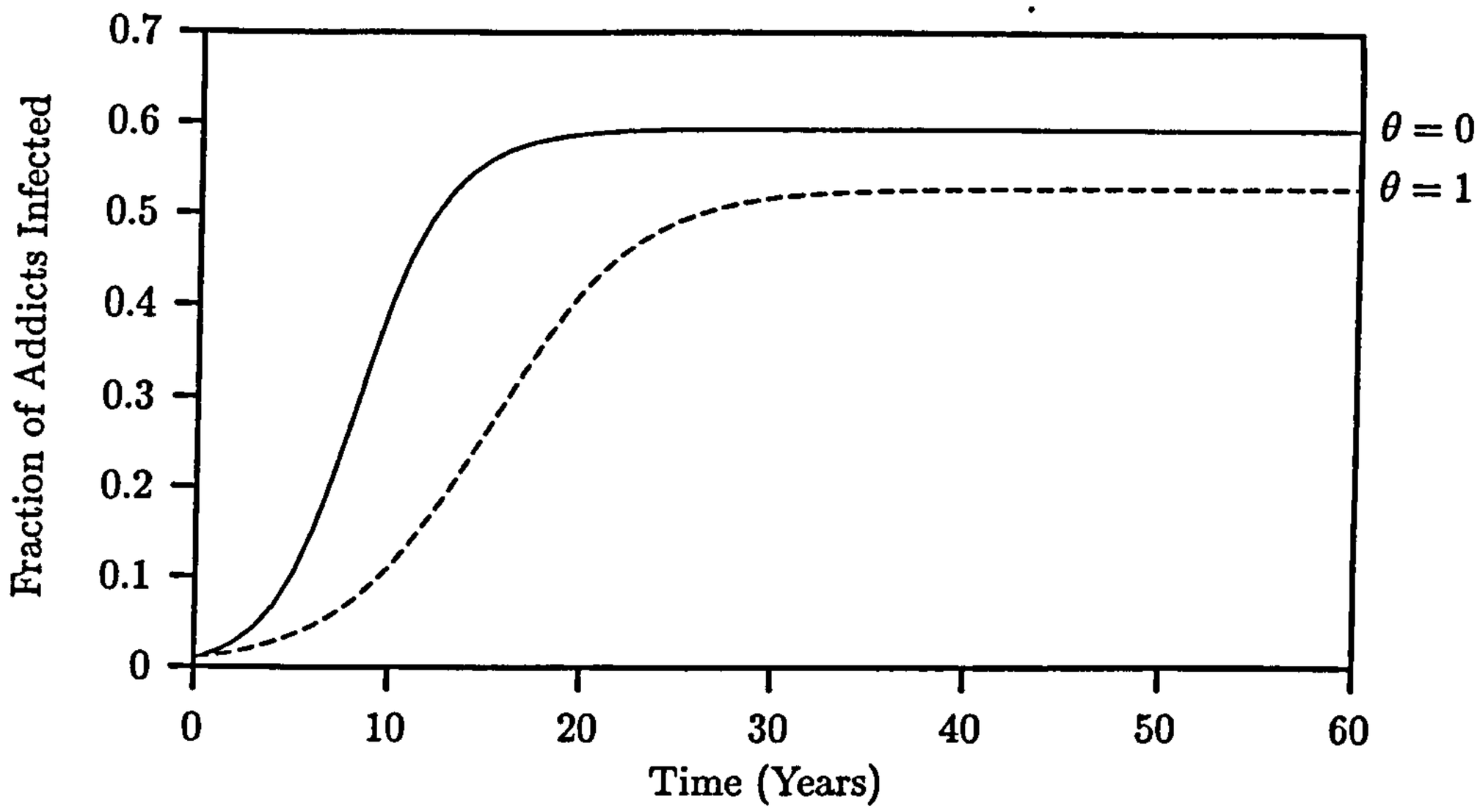
We now have a method of calibrating our models so that we ensure the cumulative viral load during the infectious lifetime of a typical addict is the same in the models being compared. However this is not yet sufficient to allow us to identify the effect of moving from single stage infectivity to three stage infectivity (in particular to three stage infectivity in both addicts and needles). The calibration in eqns in (7.5) and (7.6) corrects for differences in the viral load of the addicts between our various models, however this does not correct for differences in how addicts and needles interact in the models. If assumptions relating to how addicts and needles interact with each other cause little difference to long term prevalence levels then we can simply compare the models using calibration eqns (7.5) or (7.6) as appropriate. However if different addict-needle interaction assumptions cause a significant difference then the calibration equations on their own are not sufficient to provide a meaningful comparison. We now examine the effects caused by different addict-needle interaction assumptions in our models.

7.3 Effect of Addict-Needle Interaction Assumptions

We can consider the flushing parameter θ as representing how addicts and needles interact in the models with single stage infectious needles. In the Optimistic and Pessimistic Models the way addicts and needles interact are fixed. We now examine how the long term prevalence level of disease in addicts is affected by the choice of addict-needle interaction assumptions in the Kaplan and O'Keefe Model, the Simple Model and the Optimistic and Pessimistic Models.

Figure 7.1 shows simulations of the total fraction of infected addicts in the Kaplan and O'Keefe Model using the parameter estimates from Table 2.1 (plus $\delta = 0.101958$)

Figure 7.1: Kaplan & O'Keefe $\theta = 0$ v $\theta = 1$



except for θ, which was fixed at $\theta = 0$ and $\theta = 1$. The initial conditions in the simulations were $\pi(0) = 0.01$ and $\beta(0) = 0$. It is clear from the simulations that there is a significant difference in long term prevalence between the case where infectious needles always adopt the infectious state of the last user (full flushing) and where infectious needles remain infectious until cleaned or exchanged (no flushing). From eqns (1.1)-(1.3) we have that the long term prevalence of disease in addicts and needles are respectively

$$\pi^\star = \frac{(R_0 - 1)(\hat{\tau} + \hat{\theta})}{1 - \hat{\theta} + R_0(\hat{\tau} + \hat{\theta})}, \tag{7.7}$$

and

$$\beta^\star = \frac{1}{1 + \hat{\tau}} \left(\frac{R_0 - 1}{R_0} \right), \tag{7.8}$$

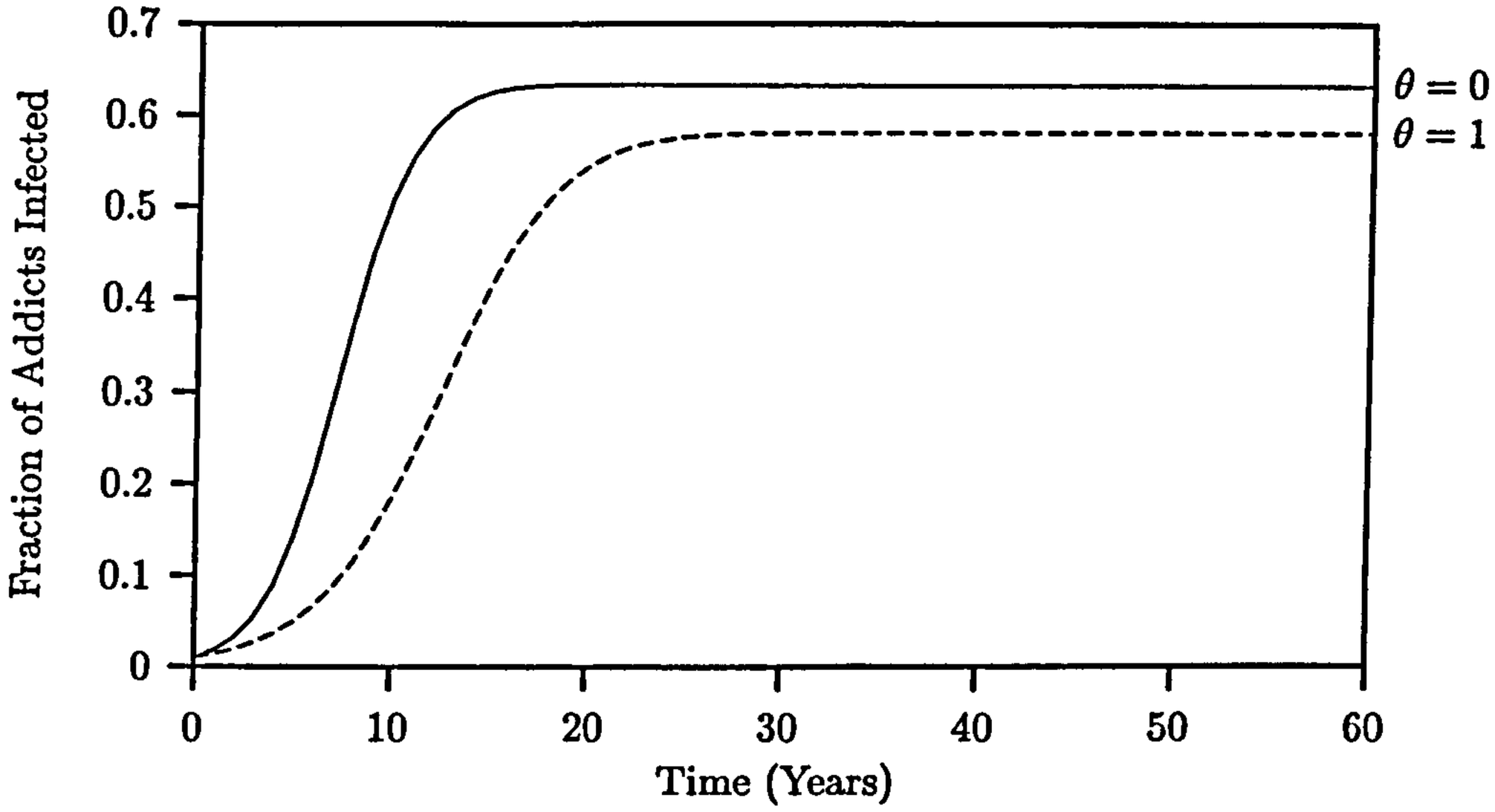
where

$$R_0 = \frac{\lambda\alpha(1 - \phi)}{(\mu + \delta)(\hat{\tau} + \hat{\theta})}. \tag{7.9}$$

Recall that $\hat{\tau} = \tau/(\lambda\gamma)$ and $\hat{\theta} = 1 - (1 - \theta)(1 - \phi)$, hence we have that $\pi^\star$ and $\beta^\star$ are both monotone decreasing in θ.

Figure 7.2 shows simulations of the total fraction of infected addicts in the Simple Model using the parameter estimates from Table 2.1 except for θ, which was fixed at $\theta = 0$ and $\theta = 1$. The initial conditions in the simulations were $\pi_1(0) = 0.01$, $\pi_2(0) = 0$, $\pi_3(0) = 0$ and $\beta(0) = 0$. As in the Kaplan and O'Keefe Model it is clear that there is a reasonable difference between the long term prevalence of disease in addicts under the assumption of full flushing and no flushing. From Chapter 2 we have that $\pi^\star$ and $\beta^\star$

Figure 7.2: Simple Model $\theta = 0$ v $\theta = 1$



are both monotone decreasing in θ since

$$\pi^\star = \frac{(R_0 - 1)(\hat{\tau} + \hat{\theta})}{1 - \hat{\theta} + R_0(\hat{\tau} + \hat{\theta})}, \tag{7.10}$$

and

$$\beta^\star = \frac{1}{1 + \hat{\tau}} \left(\frac{R_0 - 1}{R_0} \right), \tag{7.11}$$

where

$$R_0 = \frac{\lambda \alpha (1 - \phi) L}{(\mu + \delta_1)(\hat{\tau} + \hat{\theta})}, \tag{7.12}$$

and

$$L = 1 + \frac{\delta_1}{\mu + \delta_2} + \frac{\delta_1 \delta_2}{(\mu + \delta_2)(\mu + \delta_3)}. \tag{7.13}$$

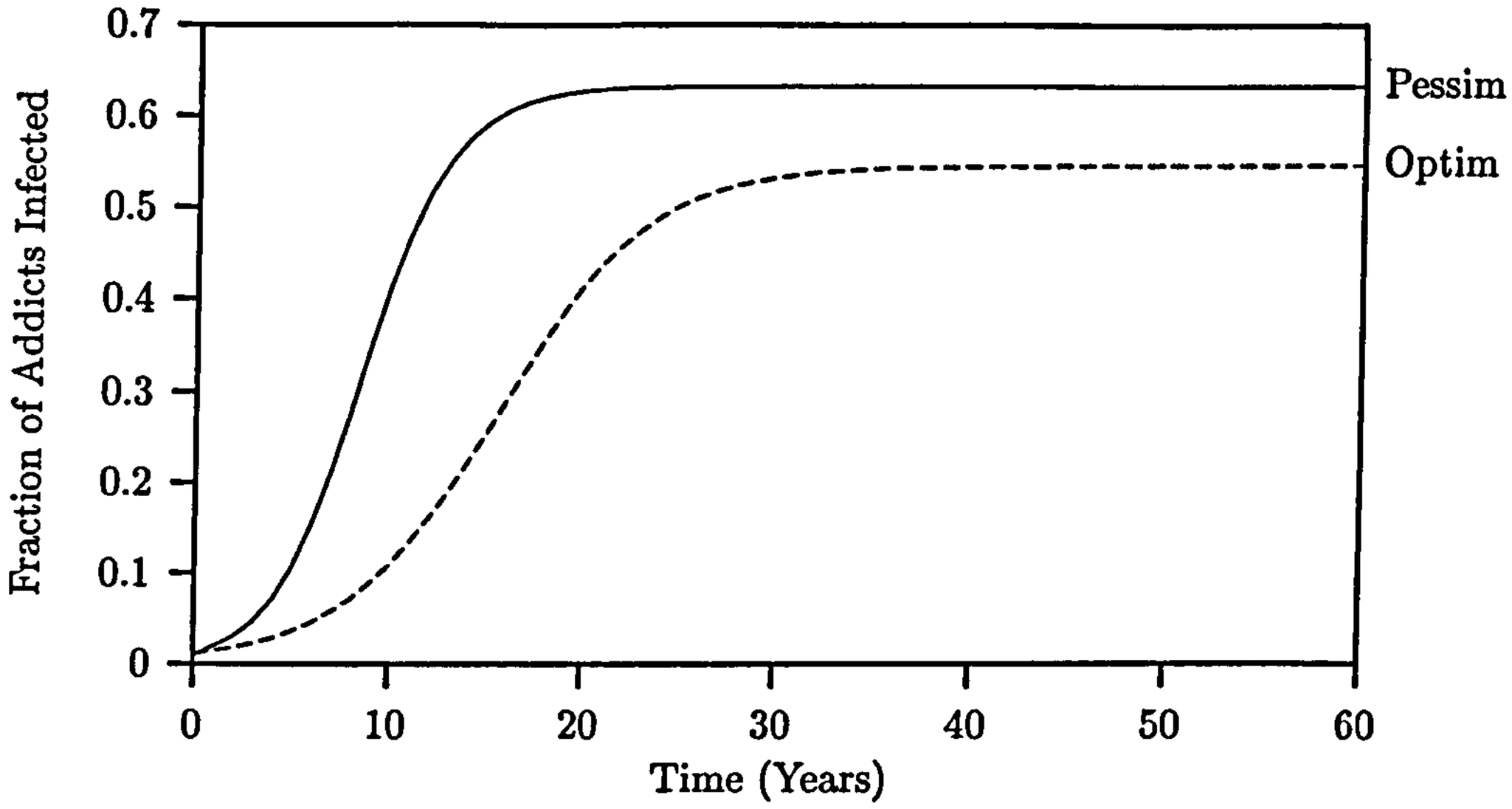
Figure 7.3 shows simulations of the total fraction of infected addicts in the Optimistic and Pessimistic Models using the parameter estimates from Table 3.1. The initial conditions in each simulation were $\pi_1(0) = 0.01$ with all other types of infectious addicts and infectious needles initially zero. The simulations clearly show that the different addict-needle interaction assumptions in these models have a considerable effect on the long term prevalence of disease in addicts. We can show analytically that the long term prevalence of disease in both addicts and needles is always higher in the Pessimistic Model than in the Optimistic Model.

Theorem 7.1 *The long term prevalence of HIV in both addicts and needles is lower in the Optimistic model than in the Pessimistic model.*

Proof.

Write $\pi_O^\star$, $\beta_O^\star$ and R_0^O to denote the values of $\pi^\star$, $\beta^\star$ and R_0 for the Optimistic Model and $\pi_P^\star$, $\beta_P^\star$ and R_0^P the corresponding values for the Pessimistic Model. From

Figure 7.3: Optimistic v Pessimistic



Theorem 4.1 we know that $\pi_P^\star$ is the unique positive solution to the equation

$$\frac{\eta_1(\mu+\delta_1)}{(1-\pi^\star)\lambda(1-\phi)} =$$

$$\frac{\alpha_1\eta_1}{\eta_1\pi^\star(1-\phi)+\hat{\tau}+\phi} + \frac{\eta_2(\hat{\tau}+\phi)\alpha_2}{(\hat{\tau}+\phi+\pi^\star(1-\phi))(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))}$$

$$+ \frac{\eta_3(\hat{\tau}+\phi)\alpha_3}{(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))(\hat{\tau}+\phi+\eta_1\pi^\star(1-\phi))}, \tag{7.14}$$

$$= (\alpha_1-\alpha_3)\frac{\eta_1}{(\eta_1\pi^\star(1-\phi)+\hat{\tau}+\phi)}$$

$$+ (\alpha_3-\alpha_2)\left\{\frac{1}{\eta_1\pi^\star(1-\phi)+\hat{\tau}+\phi}\left[\eta_1+\frac{\eta_3(\hat{\tau}+\phi)}{(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))}\right]\right\}$$

$$+ \alpha_2\left\{\frac{\eta_2(\hat{\tau}+\phi)}{(\hat{\tau}+\phi+\pi^\star(1-\phi))(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))}\right.$$

$$\left. + \frac{1}{\eta_1\pi^\star(1-\phi)+\hat{\tau}+\phi}\left[\eta_1+\frac{\eta_3(\hat{\tau}+\phi)}{(\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi))}\right]\right\}. \tag{7.15}$$

Recall that $\pi_1^\star = \pi^\star\eta_1$, $\pi_2^\star = \pi^\star\eta_2$ and $\pi_3^\star = \pi^\star\eta_3$. We have that the left hand side of eqn (7.14) is strictly increasing in $\pi^\star$ and that the right hand side is strictly decreasing in $\pi^\star$. Simplifying eqn (7.15) we find that

$$\frac{\eta_1(\mu+\delta_1)}{(1-\pi^\star)\lambda(1-\phi)} = (\alpha_1-\alpha_3)\frac{\eta_1}{\hat{\tau}+\phi+\eta_1\pi^\star(1-\phi)}$$

$$+(\alpha_3-\alpha_2)\frac{(\eta_1+\eta_3)}{\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi)}$$

$$+\alpha_2 \frac{1}{\hat{\tau} + \phi + \pi^\star(1-\phi)}. \tag{7.16}$$

We have that $\pi_O^\star = 1 - (1/R_0^O)$. Substituting this expression into the left-hand side of eqn (7.16) gives

$$\begin{aligned}
\frac{\eta_1(\mu+\delta_1)}{(1-\pi_O^\star)\lambda(1-\phi)} &= \frac{1}{L}\frac{(\mu+\delta_1)R_0^O}{\lambda(1-\phi)}, && (7.17)\\
&= \frac{1}{L(1+\hat{\tau})}\left[\alpha_1 + \frac{\alpha_2\delta_1}{\mu+\delta_2} + \frac{\alpha_3\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right], \\
&= \frac{1}{1+\hat{\tau}}\left[\alpha_1\eta_1 + \alpha_2\eta_2 + \alpha_3\eta_3\right], && (7.18)\\
&= \frac{1}{1+\hat{\tau}}\left[(\alpha_1-\alpha_3)\eta_1 + (\alpha_3-\alpha_2)(\eta_1+\eta_3) + \alpha_2\right], && (7.19)\\
&< (\alpha_1-\alpha_3)\frac{\eta_1}{\hat{\tau}+\phi+\eta_1\pi_O^\star(1-\phi)} \\
&\quad +(\alpha_3-\alpha_2)\frac{(\eta_1+\eta_3)}{\hat{\tau}+\phi+(\eta_1+\eta_3)\pi_O^\star(1-\phi)} \\
&\quad +\alpha_2\frac{1}{\hat{\tau}+\phi+\pi_O^\star(1-\phi)}. && (7.20)
\end{aligned}$$

At $\pi_O^\star$ the right hand side of eqn (7.16) exceeds the left hand side, hence it follows by the argument in the proof of Theorem 4.1 that $\pi_O^\star < \pi_P^\star$. Adding eqns (4.4)-(4.6) we have that in the Pessimistic Model

$$\beta_P^\star = \frac{\pi_P^\star}{\pi_P^\star(1-\phi)+\phi+\hat{\tau}}. \tag{7.21}$$

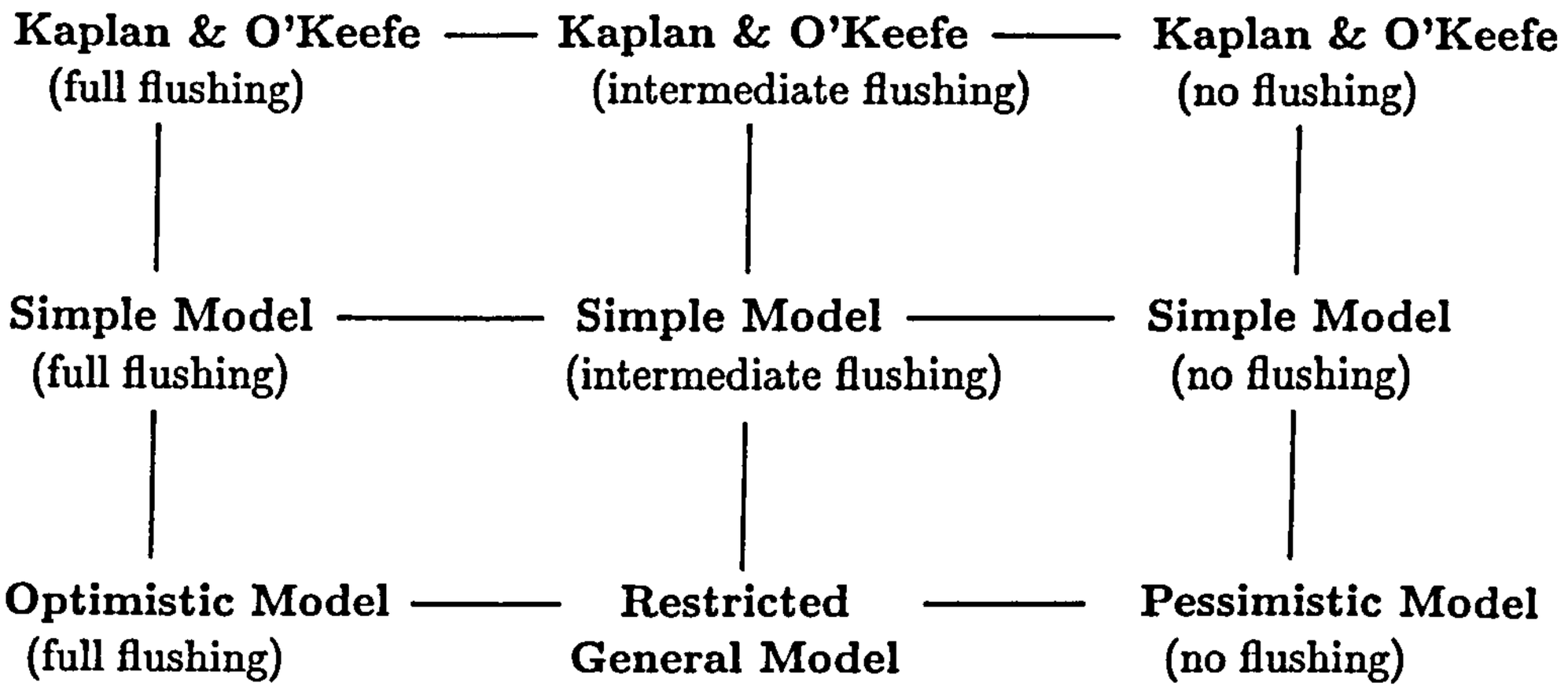
Adding eqns (3.4)-(3.6) we have that in the Optimistic Model

$$\beta_O^\star = \frac{\pi_O^\star}{1+\hat{\tau}}. \tag{7.22}$$

Hence we have directly that $\beta_O^\star < \beta_P^\star$. This completes the proof of Theorem 7.1.•

We have shown that assumptions relating to how addicts and needles interact with each other have a significant effect on the long term prevalence of disease. Hence to identify the effect of moving from single stage infectivity to three stage infectivity we must ensure that the models being compared have corresponding addict-needle interaction assumptions in addition to satisfying the appropriate calibration equation. Figure 7.4 shows graphically how the various models are related. At the top we have the original model by Kaplan and O'Keefe, this tier represents single stage infectivity models. We have split this tier into three parts according to the degree of flushing in

Figure 7.4: Model Schematic



the model. The next tier shows the stage of incorporating three types of infectious addicts (we still treat all needles as equally infectious). This tier is also split into three according to the degree of flushing in the models. The final tier represents models with both three types of infectious addicts and infectious needles. As previously discussed, intuitively we expect models with less extreme addict-needle interaction assumptions to lie between the Optimistic and Pessimistic Models. The Restricted General Model can lie anywhere on this final tier (recall that the Optimistic and Pessimistic Models are special extreme cases of the Restricted General Model). Note that we have not included the (fully) General Model in Figure 7.4 and instead defer discussion of this model until later in the chapter.

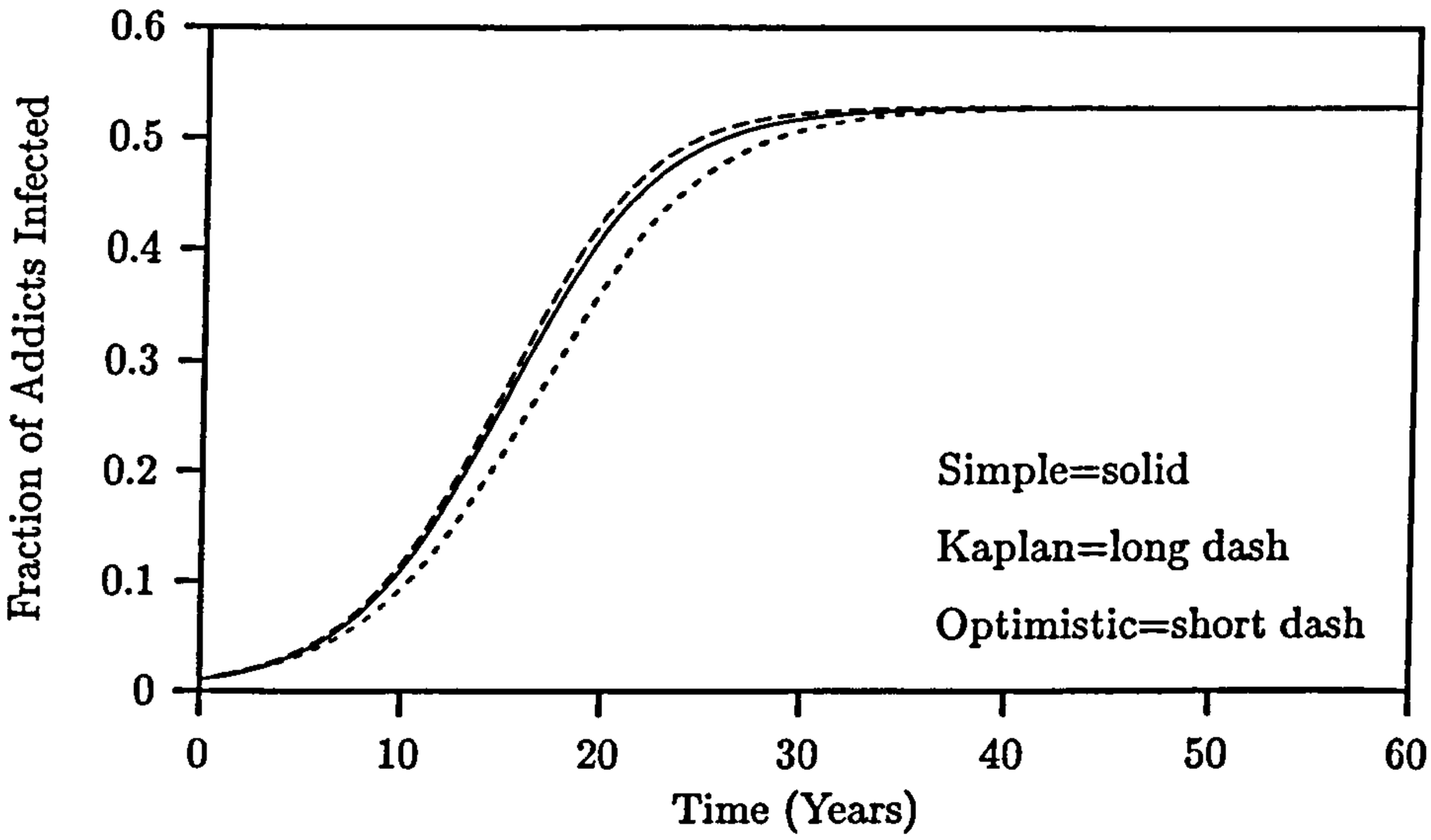
The first models we wish to compare are the Kaplan and O'Keefe with full flushing, the Simple Model with full flushing and the Optimistic Model. These models all share comparable addict-needle interaction assumptions and therefore by using the relevant calibration equations we can determine the effect of allowing addicts and needles to exist in three infectious states conditional on the assumption of full flushing.

7.4 Effect of Three Stage Infectivity with Full Flushing

Using the previous discussion on calibration we can compare the Kaplan and O'Keefe Model with the Simple Model and the Optimistic Model by first fixing all model parameters except the probability of HIV transmission (denoted by α, α' and α_2 in these

Figure 7.5: Effect of Three Stage Infectivity

Models with Full Flushing



models respectively) at their respective estimates. We then set $\alpha = 0.00601$ as in Table 2.1, and solve eqn (7.5) to find a value for α' such that the Kaplan and O'Keefe Model and Simple Model are calibrated. We then use this value of α' in eqn (7.6) and solve to find a value of α_2. Hence using $\alpha = 0.00601$ in the Kaplan and O'Keefe Model, $\alpha' = 0.005342$ in the Simple Model and $\alpha_2 = 0.002720$ in the Optimistic Model ensures that these three models can be sensibly compared. (Recall that this calibration method ignores the addict-needle interaction assumptions in the models.)

Figure 7.5 shows simulations of the total fraction of infected addicts in the Kaplan and O'Keefe Model with full flushing, the Simple Model with full flushing and the Optimistic Model, where these models are calibrated in the manner described. It is clear from the figure that there is very little difference in the behaviour of these models for the parameter estimates used in the simulations. The initial condition in each model was that a fraction 0.01 of all addicts are infectious (where all these addicts are in stage one infectivity in the three stage models). No other addicts or needles are initially infectious. These simulations suggest that the long term prevalence of HIV in addicts is unaffected by allowing addicts and needles to exist in three infectious states. However it is important to remember that we are currently assuming that needles are always left in the same infectious state as the last user (in other words needles are always flushed with probability one).

It is straightforward to show analytically that under calibration it is always true

that the long term prevalence of disease in addicts will be the same in the Kaplan and O'Keefe Model with full flushing, the Simple Model with full flushing and the Optimistic Model, (as suggested by Figure 7.5). Similarly we can show that the long term prevalence of HIV in needles is the same in each model. Using eqns (1.1)-(1.2) it is easy to show that the endemic equilibrium in the Kaplan and O'Keefe Model with full flushing is

$$(\pi^\star, \beta^\star) = \left(1 - \frac{1}{R_0^K}, \frac{1}{1+\hat{\tau}}\left[1 - \frac{1}{R_0^K}\right]\right), \tag{7.23}$$

where R_0^K is the basic reproductive number for the Kaplan and O'Keefe Model with full flushing. From Chapter 2 we have that the endemic equilibrium for the Simple Model with full flushing is

$$(\pi^\star, \beta^\star) = \left(1 - \frac{1}{R_0^S}, \frac{1}{1+\hat{\tau}}\left[1 - \frac{1}{R_0^S}\right]\right), \tag{7.24}$$

where R_0^S is the basic reproductive number for the Simple Model with full flushing. From Chapter 3 we have that the endemic equilibrium for the Optimistic Model is

$$(\pi^\star, \beta^\star) = \left(1 - \frac{1}{R_0^O}, \frac{1}{1+\hat{\tau}}\left[1 - \frac{1}{R_0^O}\right]\right), \tag{7.25}$$

where R_0^O is the basic reproductive number for the Optimistic Model. We have that

$$R_0^K = \frac{\lambda(1-\phi)\alpha}{(\mu+\delta)(1+\hat{\tau})}, \tag{7.26}$$

$$R_0^S = \frac{\lambda(1-\phi)\alpha'}{(\mu+\delta_1)(1+\hat{\tau})}\left[1 + \frac{\delta_1}{(\mu+\delta_2)} + \frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right], \tag{7.27}$$

and

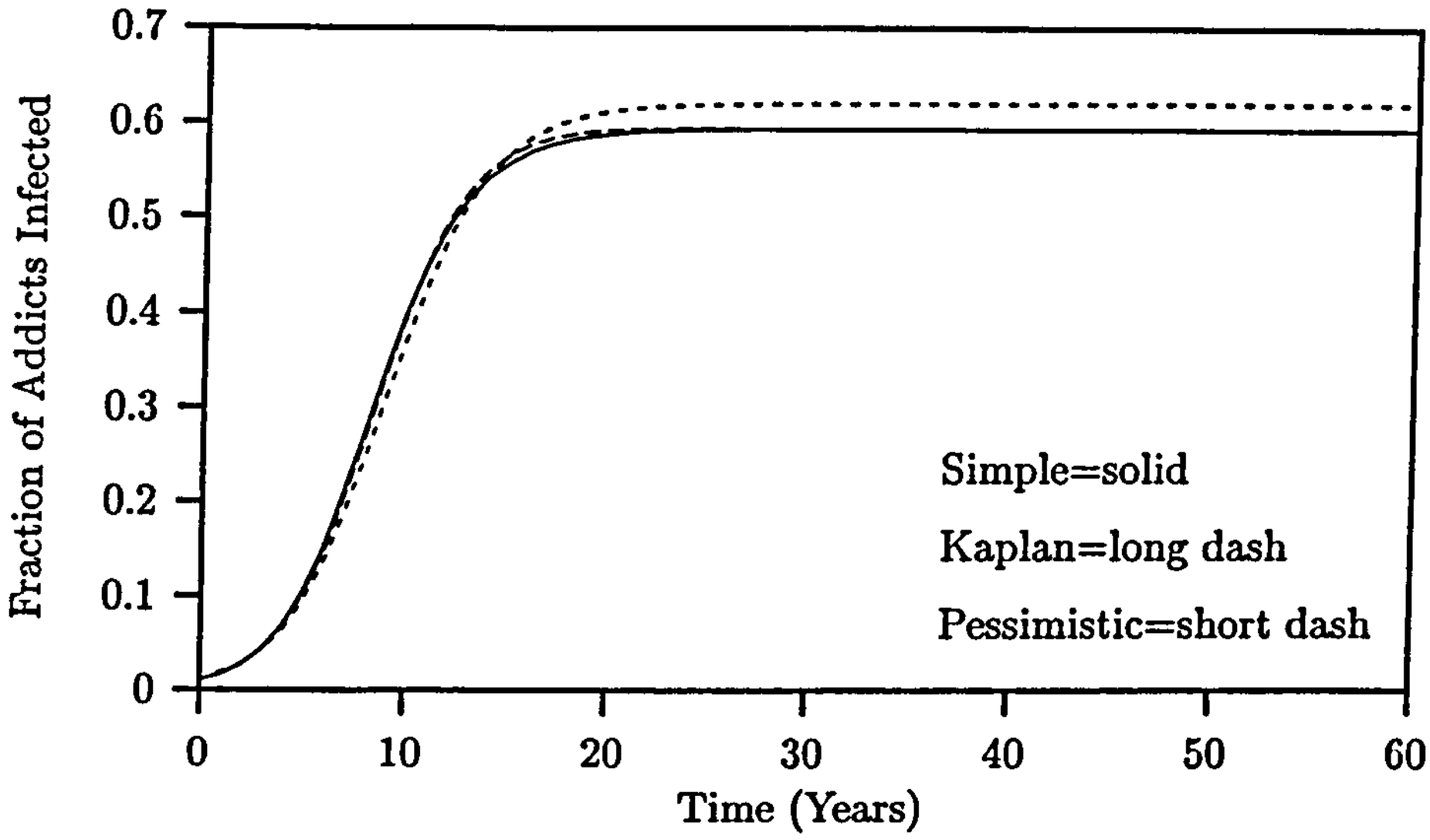
$$R_0^O = \frac{\lambda(1-\phi)\alpha_2}{(\mu+\delta_1)(1+\hat{\tau})}\left[\zeta_1 + \frac{\delta_1}{(\mu+\delta_2)} + \frac{\zeta_3\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right]. \tag{7.28}$$

Due to the way we have calibrated these models we have directly that $R_0^K = R_0^S = R_0^O$. Hence from eqns (7.23)-(7.25) it is always true that $(\pi^\star, \beta^\star)$ will be the same in the Kaplan and O'Keefe Model with full flushing, the Simple Model with full flushing and the Optimistic Model.

To summarise, we have shown that (under calibration) the long term prevalence of disease is unaffected by allowing addicts and needles to exist in three different infectious states when addicts and needles interact in accordance with the assumption of full flushing. We now examine whether this is still true for the case where addicts and needles interact in accordance with the assumption of no flushing.

Figure 7.6: Effect of Three Stage Infectivity

Models with No Flushing



7.5 Effect of Three Stage Infectivity with No Flushing

To examine the effect of moving from single stage infectivity to three stage infectivity in both addicts and needles, under the assumption that addicts and needles interact according to the assumption of no flushing, we need to compare the Kaplan and O'Keefe Model with no flushing and the Simple Model with no flushing and the Pessimistic Model. The calibration required to compare these three models is identical to that of Section 7.4. This is because as mentioned previously the calibration criteria only ensures that the cumulative viral load during the infectious lifetime of an addict is the same in the models being compared. This is independent of the assumptions relating to how addicts and needles interact with each other. Hence by using $\alpha = 0.00601$ in the Kaplan and O'Keefe Model with no flushing, $\alpha' = 0.005342$ in the Simple Model with no flushing and $\alpha_2 = 0.002720$ in the Pessimistic Model these three models are directly comparable.

Figure 7.6 shows simulations of the total fraction of infected addicts in the Kaplan and O'Keefe Model with no flushing, the Simple Model with no flushing and the Pessimistic Model where these models are calibrated in the manner described above. The initial conditions in each model were that a proportion 0.01 of addicts are initially infectious (where all these addicts are in stage one infectivity in the three stage models). No other addicts or needles are initially infectious. The simulations in Figure 7.6 show

that the long term prevalence of disease in addicts is no longer the same in each model. While we acknowledge that the trajectories of the Kaplan and O'Keefe Model and the Simple Model are not easily distinguishable in this figure the main point of the figure is to show the difference between the Pessimistic Model and the other two models. The Kaplan and O'Keefe Model with no flushing and the Simple Model with no flushing appear to have the same long term prevalence levels while the Pessimistic Model has a higher long term prevalence level.

It is straightforward to show that in general the Kaplan and O'Keefe Model with no flushing has the same long term prevalence of disease as the Simple Model with no flushing and that the Pessimistic Model has a higher long term prevalence. Using eqns (1.1)-(1.2) we have that in the Kaplan and O'Keefe Model with no flushing

$$(\pi^\star, \beta^\star) = \left(\frac{(R_0^{K\prime} - 1)(\hat{\tau} + \phi)}{1 - \phi + R_0^{K\prime}(\hat{\tau} + \phi)}, \frac{1}{1 + \hat{\tau}} \left[1 - \frac{1}{R_0^{K\prime}} \right] \right), \tag{7.29}$$

where $R_0^{K\prime}$ is the basic reproductive number for the Kaplan and O'Keefe Model with no flushing. From Chapter 2 we have that the endemic equilibrium for the Simple Model with no flushing is

$$(\pi^\star, \beta^\star) = \left(\frac{(R_0^{S\prime} - 1)(\hat{\tau} + \phi)}{1 - \phi + R_0^{S\prime}(\hat{\tau} + \phi)}, \frac{1}{1 + \hat{\tau}} \left[1 - \frac{1}{R_0^{S\prime}} \right] \right), \tag{7.30}$$

where $R_0^{S\prime}$ is the basic reproductive number for the Simple Model with no flushing. We have that

$$R_0^{K\prime} = \frac{\lambda(1-\phi)\alpha}{(\mu + \delta)(\hat{\tau} + \phi)}, \tag{7.31}$$

and

$$R_0^{S\prime} = \frac{\lambda(1-\phi)\alpha'}{(\mu + \delta_1)(\hat{\tau} + \phi)} \left[1 + \frac{\delta_1}{(\mu + \delta_2)} + \frac{\delta_1 \delta_2}{(\mu + \delta_2)(\mu + \delta_3)} \right]. \tag{7.32}$$

As in the full flushing models the way we have calibrated these models directly implies that $R_0^{K\prime} = R_0^{S\prime}$. Hence from eqns (7.29)-(7.30) it is always true that $\pi^\star$ in the Kaplan and O'Keefe Model with no flushing and the Simple Model with no flushing is the same and $\beta^\star$ in the Kaplan and O'Keefe Model with no flushing and the Simple Model with no flushing is the same. We now show that under calibration $\pi^\star$ and $\beta^\star$ are always higher in the Pessimistic Model than in the Kaplan and O'Keefe Model with no flushing or the Simple Model with no flushing.

Theorem 7.2 *Under calibration the long term prevalence of HIV in both addicts and needles is lower in the Kaplan and O'Keefe Model (with no flushing) than in the Pessimistic model.*

Proof.

Write $\pi^\star_K$ to denote the value of $\pi^\star$ for the Kaplan and O'Keefe Model with no flushing and $\pi^\star_P$ to denote the value of $\pi^\star$ in the Pessimistic Model. As in Theorem 7.1 we have that $\pi^\star$ for the Pessimistic Model is the unique positive solution to

$$\begin{aligned}\frac{\eta_1(\mu+\delta_1)}{(1-\pi^\star)\lambda(1-\phi)} &= (\alpha_1-\alpha_3)\frac{\eta_1}{\hat{\tau}+\phi+\eta_1\pi^\star(1-\phi)} \\ &\quad + (\alpha_3-\alpha_2)\frac{(\eta_1+\eta_3)}{\hat{\tau}+\phi+(\eta_1+\eta_3)\pi^\star(1-\phi)} \\ &\quad + \alpha_2\frac{1}{\hat{\tau}+\phi+\pi^\star(1-\phi)}. \end{aligned} \tag{7.33}$$

We have that the left hand side of eqn (7.33) is strictly increasing in $\pi^\star \in (0,1)$ and that the right hand side is strictly decreasing in $\pi^\star \in (0,1)$. From eqn (7.29)

$$\pi^\star_K = \frac{\frac{\lambda\alpha}{\mu+\delta} - \frac{\hat{\tau}+\phi}{1-\phi}}{1+\frac{\lambda\alpha}{\mu+\delta}}, \tag{7.34}$$

hence

$$1-\pi^\star_K = \frac{\frac{1+\hat{\tau}}{1-\phi}}{1+\frac{\lambda\alpha}{\mu+\delta}}. \tag{7.35}$$

Hence at $\pi^\star_K$ the left hand side of eqn (7.33) is

$$\frac{\eta_1\left(1+\frac{\lambda\alpha}{\mu+\delta}\right)}{\frac{\lambda}{\mu+\delta_1}(1+\hat{\tau})}. \tag{7.36}$$

The right hand side of eqn (7.33) at $\pi^\star_K$ is

$$\begin{aligned}&(\alpha_1-\alpha_3)\frac{\eta_1}{\eta_1\pi^\star_K(1-\phi)+\hat{\tau}+\phi} + (\alpha_3-\alpha_2)\frac{(\eta_1+\eta_3)}{(\eta_1+\eta_3)\pi^\star_K(1-\phi)+\hat{\tau}+\phi} \\ &\qquad + \alpha_2\frac{1}{\pi^\star_K(1-\phi)+\hat{\tau}+\phi}, \\ &\qquad > \frac{(\alpha_1-\alpha_3)\eta_1+(\alpha_3-\alpha_2)(\eta_1+\eta_3)+\alpha_2}{\hat{\tau}+\phi+\pi^\star_K(1-\phi)}, \\ &\qquad = \frac{(\alpha_1\eta_1+\alpha_2\eta_2+\alpha_3\eta_3)}{\hat{\tau}+\phi+\pi^\star_K(1-\phi)}. \end{aligned} \tag{7.37}$$

Now

$$\frac{1}{\hat{\tau}+\phi+\pi^\star_K(1-\phi)} = \frac{1+\frac{\lambda\alpha}{\mu+\delta}}{\frac{\lambda\alpha}{\mu+\delta}(1+\hat{\tau})}, \tag{7.38}$$

and from the calibration equations (7.5) and (7.6) we have that

$$\alpha_1\eta_1+\alpha_2\eta_2+\alpha_3\eta_3 = \frac{\alpha\eta_1(\mu+\delta_1)}{(\mu+\delta)}. \tag{7.39}$$

Hence the left hand side of eqn (7.33) at $\pi_K^\star$ is strictly less than the right hand side of eqn (7.33) at $\pi_K^\star$. Hence it follows that $\pi_K^\star < \pi_P^\star$. From eqn (7.29) we deduce that

$$\frac{\beta_K^\star}{1-\beta_K^\star} = \frac{\pi_K^\star}{\hat{\tau}+\phi-\pi_K^\star\phi}, \tag{7.40}$$

hence

$$\beta_K^\star = \frac{\pi_K^\star}{\hat{\tau}+\phi+\pi_K^\star(1-\phi)}. \tag{7.41}$$

As $\pi_P^\star > \pi_K^\star$ then using eqn (4.20) we have

$$\beta_P^\star = \frac{\pi_P^\star}{\hat{\tau}+\phi+\pi_P^\star(1-\phi)} > \frac{\pi_K^\star}{\hat{\tau}+\phi+\pi_K^\star(1-\phi)} = \beta_K^\star. \tag{7.42}$$

This completes the proof of Theorem 7.2. •

To summarise, we have shown that (under calibration) the long term prevalence of disease is unaffected by allowing addicts to exist in three different infectious stages (but keeping single stage infectivity in needles) when we assume that needles are never flushed. However when we additionally allow needles to exist in three different levels of infectivity and assume that addicts and needles interact according to the assumption of no flushing we find that three stage infectivity does increase the long term prevalence of disease.

7.6 The Impact of Recruitment and AIDS Mortality

We have shown that allowing addicts and needles to exist in three different infectious states does not in itself necessarily give rise to higher long term prevalence levels. Any increase in the long term prevalence of disease depends additionally on the assumptions made relating to how addicts and needles of different infectious states interact with each other.

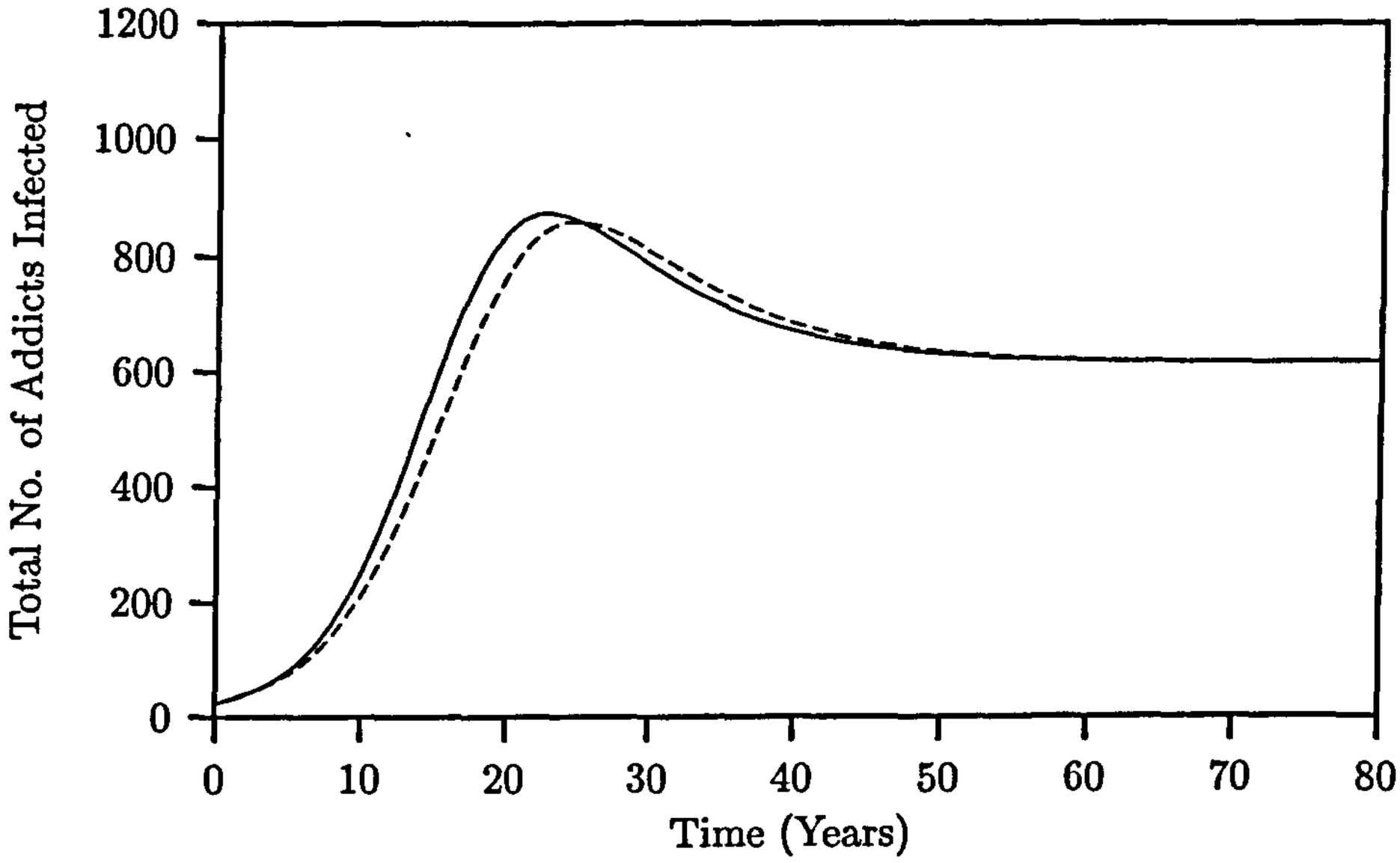
From our previous comparisons we can draw two separate conclusions. Firstly if we believe that full flushing is the most realistic approximation of how addicts and needles interact then we should conclude that three stage infectivity has no effect on the long term prevalence of disease. Alternatively if we believe that an infectious needle cannot be flushed by an uninfectious addict then we should conclude that three stage infectivity does indeed increase the long term prevalence of disease compared to single stage infectivity. However it may be premature to draw such conclusions without first taking into account the more realistic addition of addict recruitment and AIDS mortality into our assessment of three stage infectivity.

In the previous chapter we showed that the long term prevalence of disease in our three stage models is unaffected by this modification. Similarly the long term prevalence of disease in the Kaplan and O'Keefe Model is also unaffected by the inclusion of recruitment and AIDS mortality. Hence our previous comparisons and analytical results are still true when we consider the more realistic case where the population size depends on the spread of disease. Figures 7.7 and 7.8 show simulations of the total *number* of infected addicts in the Kaplan and O'Keefe Model with full flushing and the Optimistic Model, and in the Kaplan and O'Keefe Model with no flushing and the Pessimistic Model. In all these models we now include addict recruitment and AIDS mortality in a similar fashion to that modelled in the previous chapter. In each simulation it was initially assumed that 1% of addicts are infectious (and all these are in stage one infectivity in the three stage models) and the initial population size is 2,300 addicts. In the Kaplan and O'Keefe Model we use $\alpha = 0.00601$, and $\alpha_2 = 0.00272$ in the Optimistic and Pessimistic Models so that we ensure these models are suitably calibrated. Dealing first with Figure 7.7 we can see clearly that the long term number of infected addicts in these models is the same. This is not surprising since we have already shown that the equilibrium fraction of infected addicts is the same which implies that the equilibrium population size will also be the same. In Figure 7.8 we can see that the while the Pessimistic Model still has a greater long term number of infectious addicts than in the Kaplan and O'Keefe Model with no flushing the difference between these models is very small. If we had illustrated all these simulations on a single figure it would also be clear that as in Figure 6.8 in Chapter 6 the difference in the long term number of infected addicts between models with full flushing and no flushing is small.

7.7 Discussion of Comparison Results

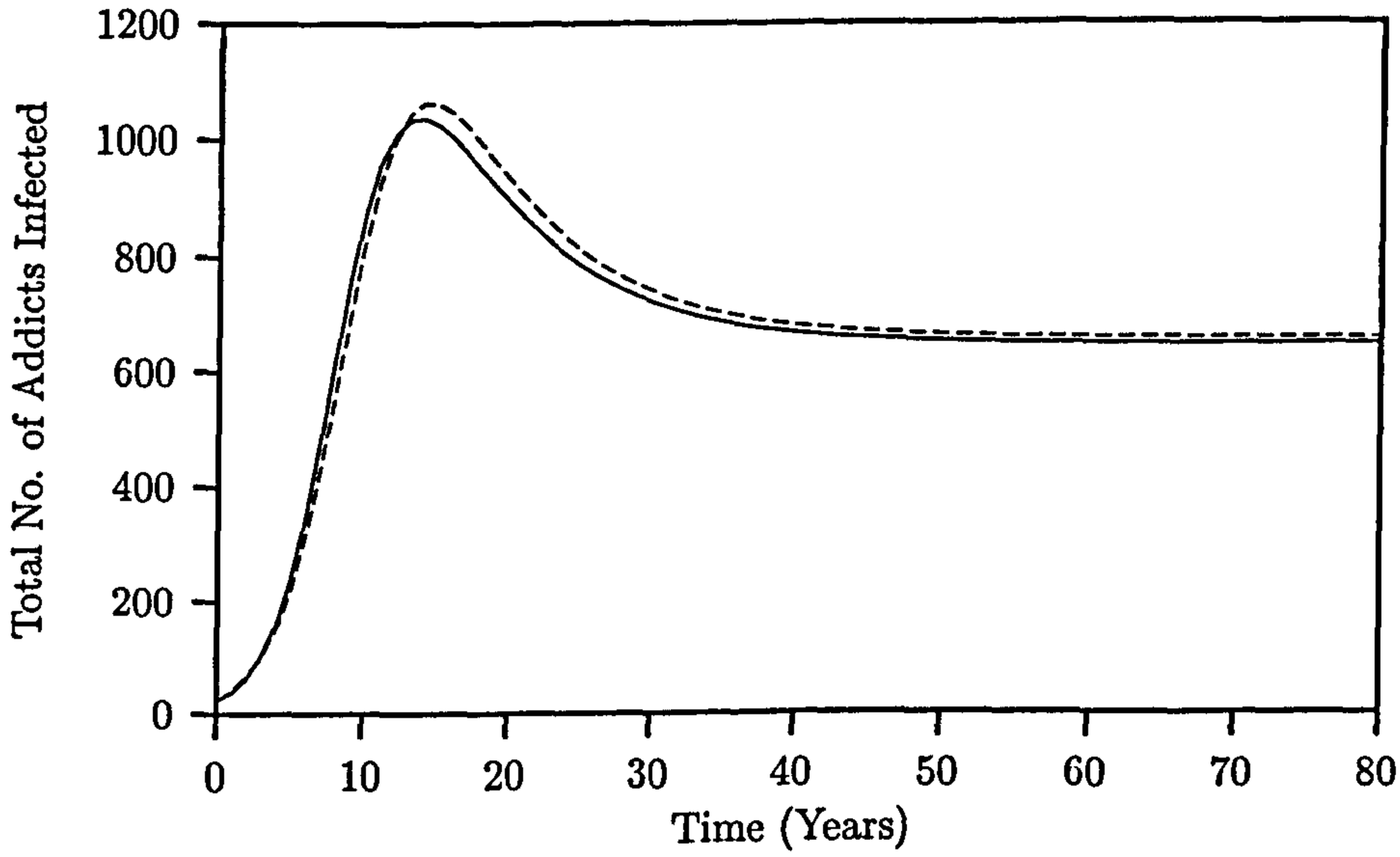
From our previous comparisons between the single stage Kaplan and O'Keefe Model and our three stage models it seems fair to conclude that moving to three stage infectivity does not cause a significant difference to the long term number of addicts infected. Analytically it is the case that the prevalence of disease is higher in the Pessimistic Model than in its single stage equivalent but in practice this increase has little effect on the long term number of addicts infected (at least for the parameter estimates used in this thesis). However there are several potential flaws in the way we have compared our various models. This deserves some discussion.

Figure 7.7: Optimistic v Kaplan and O'Keefe ($\theta = 1$)



Kaplan ($\theta = 1$) ——— Optimistic - - - - -

Figure 7.8: Pessimistic v Kaplan and O'Keefe ($\theta = 0$)



Kaplan ($\theta = 0$) ——— Pessimistic - - - - -

We have claimed that the Kaplan and O'Keefe Model with no flushing is equivalent to the Pessimistic Model. This is a reasonable assumption since (excluding cleaning) in both models an uninfectious addict cannot render a previously infectious needle virus free. However we could argue that taking the definition of flushing literally, the Pessimistic Model is only one particular variant of the class of three stage infectivity models which correspond to no flushing in single stage infectivity. For example we could argue that any selection of $p^{\star}_{ijk}$ terms in the Restricted General Model which has $p^{\star}_{i00} = 0$ for $i = 1, 2, 3$ corresponds to the Kaplan and O'Keefe Model with no flushing. The particular choice of assumptions in the Pessimistic Model satisfies this condition, however in addition this model contains the most extreme (or "pessimistic") choice of all other $p^{\star}_{ijk}$ terms. Hence we expect that any other special case of the Restricted General Model which has $p^{\star}_{i00} = 0$ for $i = 1, 2, 3$ will have a lower long term level of disease than in the Pessimistic Model. Therefore it may be the case that a three stage model which has $p^{\star}_{i00} = 0$ for $i = 1, 2, 3$ gives rise to a lower long term level of disease than in the Kaplan and O'Keefe Model with no flushing. This comparison problem does not affect the Optimistic Model since full flushing (the assumption that a needle always adopts the infectious state of the last user) uniquely identifies a set of $p^{\star}_{ijk}$ terms in three stage infectivity.

There is a second related problem with our previous comparisons. So far we have interpreted the flushing parameter θ in single stage models as the probability that an infectious needle is left uninfectious after use by an uninfectious addict. Up to now we have taken this definition literally, for example we treated $\theta = 0$ as equivalent to $p^{\star}_{100} = p^{\star}_{200} = p^{\star}_{300} = 0$. However an alternative interpretation of $\theta = 0$ is the probability that a needle with viral load proportional to α cannot have its viral load reduced to zero after use by an uninfectious addict. However in our three stage infectivity models we now have three types of needles. The lowest infectivity needle has a viral load proportional to α_2 which will be lower than the viral load of an infectious needle in a single stage model. Hence it may be that a needle in stage two infectivity can have its viral load reduced to zero and this is still equivalent to $\theta = 0$ in single stage models. Similar interpretational problems also exist with stage one and stage three needles. Therefore it may be the case that perhaps we cannot match $\theta = 0$ or indeed any value of θ with equivalent addict-needle interaction assumptions in three stage infectivity.

Given the above discussion it seems plausible that we cannot match the flushing probability in single stage infectivity with corresponding addict-needle interaction as-

sumptions in three stage infectivity. However our previous comparisons still represent a logical basis for assessing the effect of three stage infectivity. If we cannot match θ and $p^{\star}_{ijk}$ terms then an alternative is to compare upper and lower bound models in single stage infectivity with upper and lower bound models in three stage infectivity. In this way we can determine whether existing best and worst case scenarios based on the assumption of constant (single stage) infectivity require re-assessing in light of variable (three stage) infectivity. It makes sense to examine both the best and worst which may occur in order to provide balanced information which can then be used as a basis for determining future levels of health provision and intervention.

Since the endemic equilibrium prevalence of disease in addicts in the Kaplan and O'Keefe Model is monotone decreasing in θ then using $\theta = 1$ gives a lower bound for the spread of disease and similarly $\theta = 0$ gives an upper bound. Due to the addict-needle interaction assumptions in the Optimistic Model we intuitively expect this model to represent a lower bound for the prevalence of disease among addicts in three stage infectivity and similarly we expect the Pessimistic Model to represent an intuitive upper bound for the prevalence of disease among addicts. Therefore comparing lower bound single stage models with lower bound three stage models, and upper bound single stage models with upper bound three stage models, obviously gives us exactly the same comparisons as before (but justified in a different way). Hence our previous conclusion that moving from single stage infectivity to three stage infectivity has little effect on the long term number of infectious addicts still stands. As a final comment before we discuss the General Model it is of interest to ask whether we could have reasonably expected this result.

Dietz et al (1993) discuss a model of the spread of HIV and examine the impact of moving from constant to variable infectivity where HIV is transmitted sexually. In particular they include in their model the formation and separation of pairs of heterosexuals. Andersson and Britton (1998) examine the effect of heterogeneity using a more general model which does not focus on HIV and groups the population into household units rather than partnerships.

Dietz et al (1993) find that under suitable calibration R_0 is less when infectivity is considered variable as opposed to constant. They illustrate that introducing variability into infectivity can decrease the infection probability per partner. Intuitively a reason for R_0 being less is that the partnerships considered by Dietz are relatively long compared to the infectious periods of an individual. Therefore while an individual may be

highly infectious he or she may be in a long term partnership and as such his or her high level of infectivity does not produce a signficant number of new infections.

Andersson and Britton (1998) also discuss the impact of introducing heterogeneity into a homogeneous population. This paper focuses on quantifying the statement "if the disease is very contagious then homogenizing the population increases the size of the epidemic, while for a less infectious disease the largest epidemic arises in a heterogeneous setup". They discuss a number of models and examine the spread of disease among household units and among the population as a whole. It seems reasonable to consider the grouping of the population into households as an extension of the pair formation case discussed by Dietz et al (1993).

While both Dietz et al (1993) and Andersson and Britton (1998), put forward some intuitive arguments for assessing the impact of moving from constant to variable infectivity it is not clear how applicable these arguments are when applied to populations of both addicts and needles. The vector element in our models complicates such matters. However these papers do suggest that the limited effect of variable infectivity in our models may be because the length of an addict-needle partnership is extremely short, and therefore the disease spreads in a similar fashion under both constant and variable infectivity. In Chapter 11 we mention that the spread of HIV when addict-needle partnerships are longer, such as sharing needles in "social networks", is an area for future study.

7.8 General Model and Single Stage Infectivity

Up to now in this chapter we have avoided discussion of the General Model. This model contains a completely general addict-needle interaction structure, in particular we have no longer placed any restrictions on interactions between uninfectious needles and infectious addicts. However this increased flexibility (over the Restricted General Model) causes problems in comparing this model with the Kaplan and O'Keefe Model. The crux of the problem is that Kaplan and O'Keefe's Model (and Kaplan's original model) assumes that an uninfectious needle is always left infectious after use by an infectious addict. We have made an analogous assumption in the Simple Model and in the Restricted General Model (since in the latter we have assumed that $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$), however we have dropped this arguably extreme assumption in the General Model. This means that the General Model does not fit into a comparable

framework with the Kaplan and O'Keefe Model (such as that illustrated in Figure 7.4) since in the former, infectious addicts can leave previously uninfectious needles still uninfectious, whereas in the latter the same needle is always left infectious. To produce a fair comparison between the General Model and the Kaplan and O'Keefe Model we first need to extend the latter single stage model to also allow infectious addicts to leave previously uninfectious needles uninfectious after use. This is an analogous extension to that of moving from the Restricted General Model to the General Model.

It seems intuitive that at least some (if not all) of the time an infectious addict will leave infectious material behind in a needle after use. Hence given the choice between classifying a previously uninfectious needle as infectious or uninfectious after use by an infectious addict, it seems only prudent to classify this needle as infectious. A similar argument can be used to justify $\theta = 0$ since again it seems intuitive that at least some infectious material will remain behind in a previously infectious needle after use by an uninfectious addict. Hence taking a cautious approach there is good reason for assuming as Kaplan and O'Keefe do that previously uninfectious needles always become infectious after use by infectious addicts and previously infectious needles always remain infectious after use by uninfectious addicts. However it is clear that in both these interactions the assumptions made are more pessimistic than would be realistically expected to occur, but in the absence of hard data (which is almost impossible to obtain) such a cautious approach seems sensible. We can consider this pessimism necessary as a result of the "broad brush" approach to modelling addict-needle interactions.

Ideally a single stage infectivity model should have two flushing parameters, ψ and θ, where ψ denotes the probability that an uninfectious needle is left infectious after use by an infectious addict. Greenhalgh and Hay (1997) extend Kaplan's original model to allow for just such an interaction structure. The Kaplan and O'Keefe Model can easily be extended to include both types of flushing, this gives us the following system of equations:

$$\frac{d\pi}{dt} = (1-\pi)\lambda\alpha(1-\phi)\beta - (\mu+\delta)\pi, \tag{7.43}$$

and

$$\frac{d\beta}{dt} = (1-\beta)\lambda\gamma\pi\psi - (1-\pi)\beta\lambda\gamma\hat{\theta} - \beta\tau. \tag{7.44}$$

We now have a single stage model which fits into a framework similar to that in the General Model. However the problem remains (as discussed above) of how to choose a set of $p^{\star}_{ijk}$ terms which correspond to any particular set of θ and ψ values (or vice

versa). There are at least two intuitive ways of doing this which we now discuss.

7.8.1 Calibration Method One

There are three main differences between the General Model and the model in eqns (7.43)-(7.44). Firstly the cumulative viral load during an addict's lifetime may be different. Secondly the interaction assumptions between infectious addicts and uninfectious needles are dealt with differently, as are the interaction assumptions between uninfectious addicts and infectious needles. As already discussed the first of these differences can be dealt with using the calibration eqns (7.5) and (7.6). We now deal with the second difference; how to choose equivalent values for ψ and the interaction terms $p^{\star}_{0jk}$ for $j,k = 1,2,3$. One possibility is that we should require that the average total amount of virus transmitted by a single infectious addict to needles at the disease-free equilibrium should be the same in both our models. This suggests the following calibration equation:

$$\begin{aligned}\frac{\lambda\psi\alpha}{\mu+\delta} &= \frac{\lambda}{\mu+\delta_1}(\alpha_1 p^{\star}_{011} + \alpha_2 p^{\star}_{012} + \alpha_3 p^{\star}_{013}) + \frac{\lambda\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}\alpha_2 p^{\star}_{022} \\ &\quad + \frac{\lambda\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}(\alpha_2 p^{\star}_{032} + \alpha_3 p^{\star}_{033}). \qquad (7.45)\end{aligned}$$

Clearly $\psi \geq 0$ for any choice of $p^{\star}_{ijk}$ terms and we have that

$$\begin{aligned}\frac{\lambda\psi\alpha}{\mu+\delta} &\leq \frac{\lambda}{\mu+\delta_1}\alpha_1 + \frac{\lambda\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}\alpha_2 + \frac{\lambda\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\alpha_3, \\ &= \frac{\lambda\alpha}{\mu+\delta}, \qquad \text{using eqns (7.5) and (7.6).}\end{aligned}$$

Therefore we have that $0 \leq \psi \leq 1$ as is sensible since ψ is a probability. The above calibration implies that for a single stage model to be equivalent to the Restricted General Model (where $p^{\star}_{011} = p^{\star}_{022} = p^{\star}_{033} = 1$) it must have $\psi = 1$. This is what we would expect and provides us with some confidence in the suitability of this calibration method. Note that we use $p^{\star}_{ijk}$ terms in our calibration rather than p_{ijk} terms, this is natural since ψ and θ do not include the effect of cleaning.

The final difference between our models is similar to that just discussed but for interactions between infectious needles and uninfectious addicts. We can choose equivalent values for θ and the interaction terms $p^{\star}_{i0k}$ for $i,k = 1,2,3$ by requiring that the average total amount of virus removed from all secondary infectious needles on their first use by a susceptible addict at the disease-free equilibrium should be the same in

both our models. This suggests the following calibration equation:

$$\begin{aligned}\frac{\lambda\psi\alpha\theta}{\mu+\delta} &= \frac{\lambda}{(\mu+\delta_1)}p^{\star}_{011}\left(\alpha_1 p^{\star}_{100}+(\alpha_1-\alpha_2)p^{\star}_{102}+(\alpha_1-\alpha_3)p^{\star}_{103}\right)\\ &\quad+\frac{\lambda}{(\mu+\delta_1)}\left(p^{\star}_{012}+\frac{\delta_1}{\mu+\delta_2}p^{\star}_{022}+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}p^{\star}_{032}\right)\alpha_2 p^{\star}_{200}\\ &\quad+\frac{\lambda}{(\mu+\delta_1)}\left(p^{\star}_{013}+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}p^{\star}_{033}\right)(\alpha_3 p^{\star}_{300}+(\alpha_3-\alpha_2)p^{\star}_{302}). \quad (7.46)\end{aligned}$$

For example the average number of needles left in state one infectivity during the infectious lifetime of an addict is $(\lambda p^{\star}_{011})/(\mu+\delta_1)$, and each of these needles has a viral load proportional to α_1. However the amount of virus in each of these needles can be reduced due to use by an uninfectious addict, therefore the average total amount of virus removed at the first injection from all these state one needles by uninfectious addicts is

$$\frac{\lambda}{(\mu+\delta_1)}p^{\star}_{011}\left(\alpha_1 p^{\star}_{100}+(\alpha_1-\alpha_2)p^{\star}_{102}+(\alpha_1-\alpha_3)p^{\star}_{103}\right).$$

The other terms in eqn (7.46) follow similarly but are for the amount of virus contained in state two and state three infectious needles. From eqn (7.46) it is clear that $\theta \geq 0$, we also have that

$$\begin{aligned}\frac{\lambda\psi\alpha\theta}{\mu+\delta} &\leq \frac{\lambda}{(\mu+\delta_1)}p^{\star}_{011}\alpha_1+\frac{\lambda}{(\mu+\delta_1)}\left(p^{\star}_{012}+\frac{\delta_1}{\mu+\delta_2}p^{\star}_{022}+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}p^{\star}_{032}\right)\alpha_2\\ &\quad+\frac{\lambda}{(\mu+\delta_1)}\left(p^{\star}_{013}+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}p^{\star}_{033}\right)\alpha_3,\\ &= \frac{\lambda\psi\alpha}{\mu+\delta}, \qquad \text{using eqn (7.45).}\end{aligned}$$

Therefore we must have that $0 \leq \theta \leq 1$ which again is sensible since θ is a probability. Since the Optimistic and Pessimistic Models are special cases of the Restricted General Model we have from eqn (7.45) that $\psi = 1$ in both these models. Substituting into eqn (7.46) the $p^{\star}_{ijk}$ terms for the Optimistic Model we find that

$$\frac{\lambda\alpha\theta}{\mu+\delta} = \frac{\lambda}{\mu+\delta_1}\alpha_1+\frac{\lambda\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}\alpha_2+\frac{\lambda\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\alpha_3,$$

which implies that $\theta = 1$ since we require that eqns (7.5) and (7.6) are satisfied. Similarly for the Pessimistic Model we find that $\lambda\alpha\theta/(\mu+\delta) = 0$, which implies that $\theta = 0$. As before these results are sensible and provide confidence in our calibration method.

Hence to calibrate the General Model and the Kaplan and O'Keefe Model (with $0 \leq \psi \leq 1$ and $0 \leq \theta \leq 1$) and therefore quantify the effect of three stage infectivity

we do the following: First we estimate all the parameters in each of our models with the exception of ψ and θ in the single stage model and α_1, α_2 and α_3 in the General Model. We then use eqns (7.5) and (7.6) to find the appropriate value of α_2 (which also gives α_1 and α_3) in the General Model. We next use eqn (7.45) to find the value of ψ which corresponds to the current set of $p^\star_{ijk}$ terms in the General Model. Finally having found ψ we can then use eqn (7.46) to find the corresponding value of θ.

We now illustrate an example of this calibration using the sets of $p^\star_{ijk}$ terms in Table 7.1. These $p^\star_{ijk}$ values have been deliberately chosen to be fairly moderate in that the final state of the needle is not particularly biased towards either its state prior to use or that of the current user. Using eqns (7.5) and (7.6) we again require that $\alpha_2 = 0.00272$, $\alpha_1 = 0.0136$ and $\alpha_3 = 0.00816$. Using the $p^\star_{0jk}$ terms in Table 7.1 together with eqn (7.45) we find that sets A2-C2 correspond to $\psi = 0.52, 0.42$ and 0.53 respectively. Using these values of ψ together with eqn (7.46) we additionally find that sets A2-C2 also correspond to $\theta = 0.78, 0.53$ and 0.26 respectively.

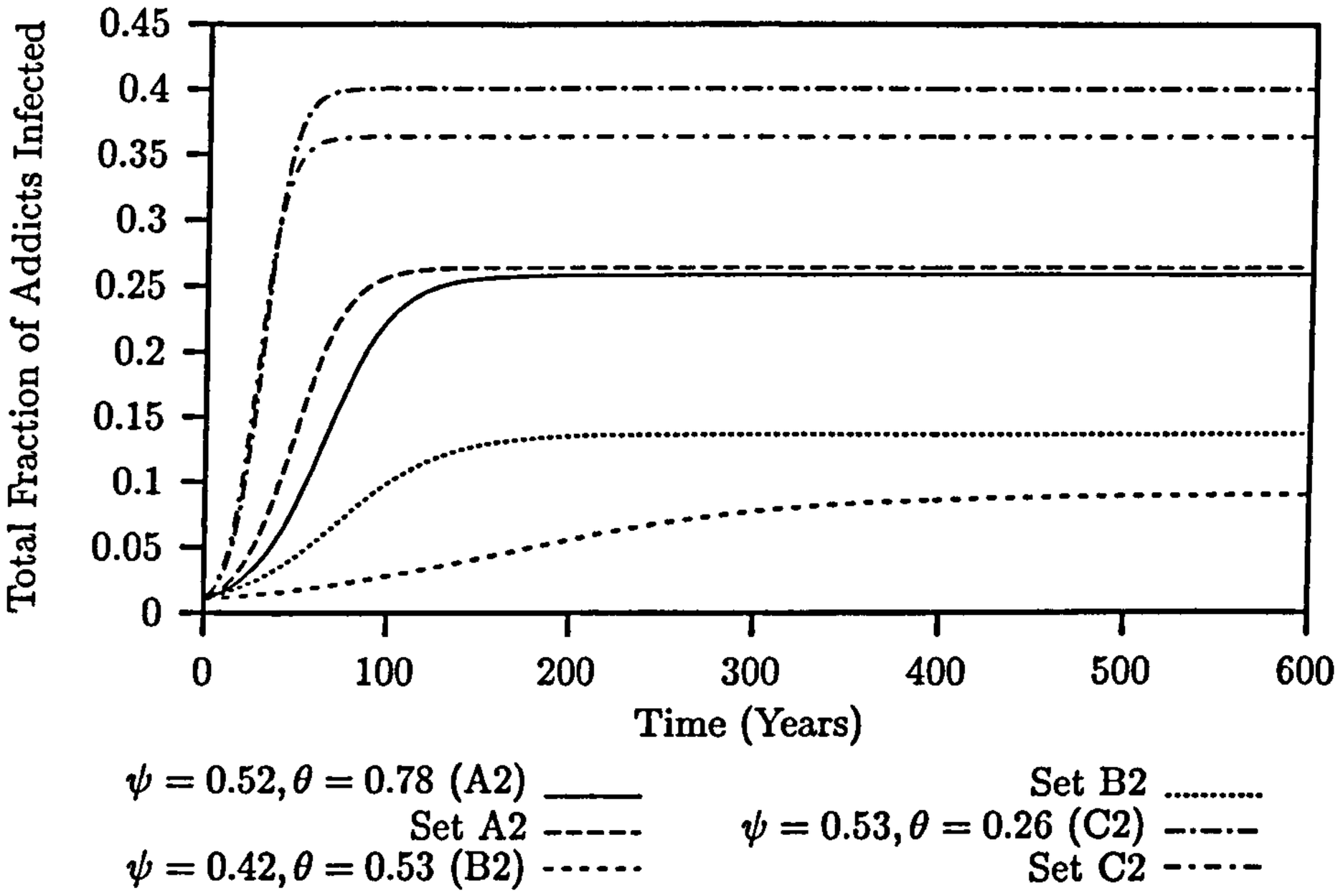
An additional complication of comparing single stage and three stage models is that in three stage infectivity we have interactions between addicts who and needles which are all infectious but are currently in different infectious classes. This does not occur in single stage infectivity since we only have a single class of infectious needle and a single class of infectious addict. Hence we can use any values for the $p^\star_{ijk}$ terms where $i, j, k = 1, 2, 3$ (such that the inequalities (5.1)-(5.10) are satisfied) and still have that the General Model and the Kaplan and O'Keefe Model (with $0 \leq \theta \leq 1$ and $0 \leq \psi \leq 1$) are correctly calibrated.

Figure 7.9 shows simulations of the total fraction of infected addicts using the Kaplan and O'Keefe Model for the values of ψ and θ mentioned above and the General Model using $p^\star_{ijk}$ sets A2-C2 in Table 7.1. It was initially assumed that 1% of addicts are in stage one infectivity and no other addicts or needles are infectious. It is not clear from these simulations whether moving to three stage infectivity either increases or decreases the long term prevalence of disease. For example the single stage model corresponding to set C2 suggests a higher level of disease while the model corresponding to set B2 has a lower level of disease. As mentioned above an additional complication in assessing the effect of three stage infectivity is that a different choice of $p^\star_{ijk}$ terms (where $i, j, k = 1, 2, 3$) may cause an increase or decrease in the long term prevalence illustrated. In any case our simulations suggest that a simple summary of the effect of three stage infectivity is not appropriate, and that this effect depends on the particular

Table 7.1: Addict-Needle Assumptions (No Cleaning)

$p^{\star}_{ijk}$	A2	B2	C2
$p^{\star}_{010}$	0.0	0.25	0.0
$p^{\star}_{011}$	0.05	0.25	0.3
$p^{\star}_{012}$	0.3	0.25	0.5
$p^{\star}_{013}$	0.65	0.25	0.2
$p^{\star}_{020}$	0.5	0.5	0.5
$p^{\star}_{022}$	0.5	0.5	0.5
$p^{\star}_{030}$	0.1	0.33	0.1
$p^{\star}_{032}$	0.4	0.33	0.4
$p^{\star}_{033}$	0.5	0.33	0.5
$p^{\star}_{100}$	0.7	0.25	0.05
$p^{\star}_{101}$	0.05	0.25	0.7
$p^{\star}_{102}$	0.2	0.25	0.05
$p^{\star}_{103}$	0.05	0.25	0.2
$p^{\star}_{121}$	0.1	0.33	0.7
$p^{\star}_{122}$	0.7	0.33	0.1
$p^{\star}_{123}$	0.2	0.33	0.2
$p^{\star}_{131}$	0.3	0.5	0.7
$p^{\star}_{133}$	0.7	0.5	0.3
$p^{\star}_{200}$	0.7	0.5	0.3
$p^{\star}_{202}$	0.3	0.5	0.7
$p^{\star}_{211}$	0.7	0.33	0.3
$p^{\star}_{212}$	0.1	0.33	0.3
$p^{\star}_{213}$	0.2	0.33	0.4
$p^{\star}_{232}$	0.3	0.5	0.5
$p^{\star}_{233}$	0.7	0.5	0.5
$p^{\star}_{300}$	0.7	0.33	0.1
$p^{\star}_{302}$	0.2	0.33	0.2
$p^{\star}_{303}$	0.1	0.33	0.7
$p^{\star}_{311}$	0.7	0.5	0.3
$p^{\star}_{313}$	0.3	0.5	0.7
$p^{\star}_{322}$	0.7	0.5	0.3
$p^{\star}_{323}$	0.3	0.5	0.7

Figure 7.9: Single Stage v Three Stage (Fully General) : Calibration Method One



addict-needle interaction assumptions in our models.

7.8.2 Calibration Method Two

We now discuss another intuitive criteria for calibrating our three stage and single stage models. We have shown previously that the basic reproductive number is fundamental in determining the behaviour of both single stage and three stage models. Therefore a natural calibration criteria is to ensure that R_0 is the same in both types of models. For the Kaplan and O'Keefe Model with $0 \leq \psi \leq 1$ we have that

$$R_0^K = \frac{\lambda \psi \alpha}{(\mu + \delta)\theta}, \tag{7.47}$$

when interventions such as needle cleaning and needle exchange do not occur. It seems sensible to exclude external interventions such as these from the calibration method as we wish to calibrate the models on purely biological grounds. In the General Model we have that

$$\begin{aligned} R_0^{GEN} &= \frac{\lambda p^\star_{011}}{(\mu+\delta_1)}\left\{\frac{\alpha_1}{1-p^\star_{101}} + \frac{p^\star_{103}}{1-p^\star_{101}}\left[\frac{\alpha_3}{1-p^\star_{303}} + \frac{p^\star_{302}}{(1-p^\star_{303})}\frac{\alpha_2}{(1-p^\star_{202})}\right]\right. \\ &+ \left.\frac{p^\star_{102}}{(1-p^\star_{101})}\frac{\alpha_2}{(1-p^\star_{202})}\right\} \\ &+ \frac{\lambda}{(\mu+\delta_1)}\left[p^\star_{012} + \frac{\delta_1 p^\star_{022}}{(\mu+\delta_2)} + \frac{\delta_1\delta_2 p^\star_{032}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\frac{\alpha_2}{(1-p^\star_{202})} \end{aligned}$$

$$+ \quad \frac{\lambda}{(\mu+\delta_1)}\left[p^{\star}_{013}+\frac{\delta_1\delta_2 p^{\star}_{033}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\left[\frac{\alpha_3}{1-p^{\star}_{303}}+\frac{p^{\star}_{302}}{(1-p^{\star}_{303})}\frac{\alpha_2}{(1-p^{\star}_{202})}\right], \tag{7.48}$$

in the absence of needle exchange and needle cleaning. Therefore

$$\begin{aligned} R_0^{GEN} &\geq \frac{\lambda p^{\star}_{011}}{(\mu+\delta_1)}\alpha_1+\frac{\lambda}{(\mu+\delta_1)}\left[p^{\star}_{012}+\frac{\delta_1 p^{\star}_{022}}{(\mu+\delta_2)}+\frac{\delta_1\delta_2 p^{\star}_{032}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\alpha_2 \\ &+ \quad \frac{\lambda}{(\mu+\delta_1)}\left[p^{\star}_{013}+\frac{\delta_1\delta_2 p^{\star}_{033}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\alpha_3, \\ &= \frac{\lambda\psi\alpha}{\mu+\delta}, \qquad \text{using eqn (7.45)}, \\ &= R_0^K\theta. \end{aligned}$$

Therefore again we have that $0 \leq \theta \leq 1$ as should be the case. Equating R_0^K and R_0^{GEN} and substituting in the $p^{\star}_{ijk}$ values corresponding to the Pessimistic Model implies that $\theta = 0$. Similarly using the interaction assumptions in the Optimistic Model requires that

$$\frac{\lambda\psi\alpha}{(\mu+\delta)\theta} = \frac{\lambda}{\mu+\delta_1}\left[\alpha_1+\frac{\delta_1\alpha_2}{\mu+\delta_2}+\frac{\delta_1\delta_2\alpha_3}{(\mu+\delta_2)(\mu+\delta_3)}\right].$$

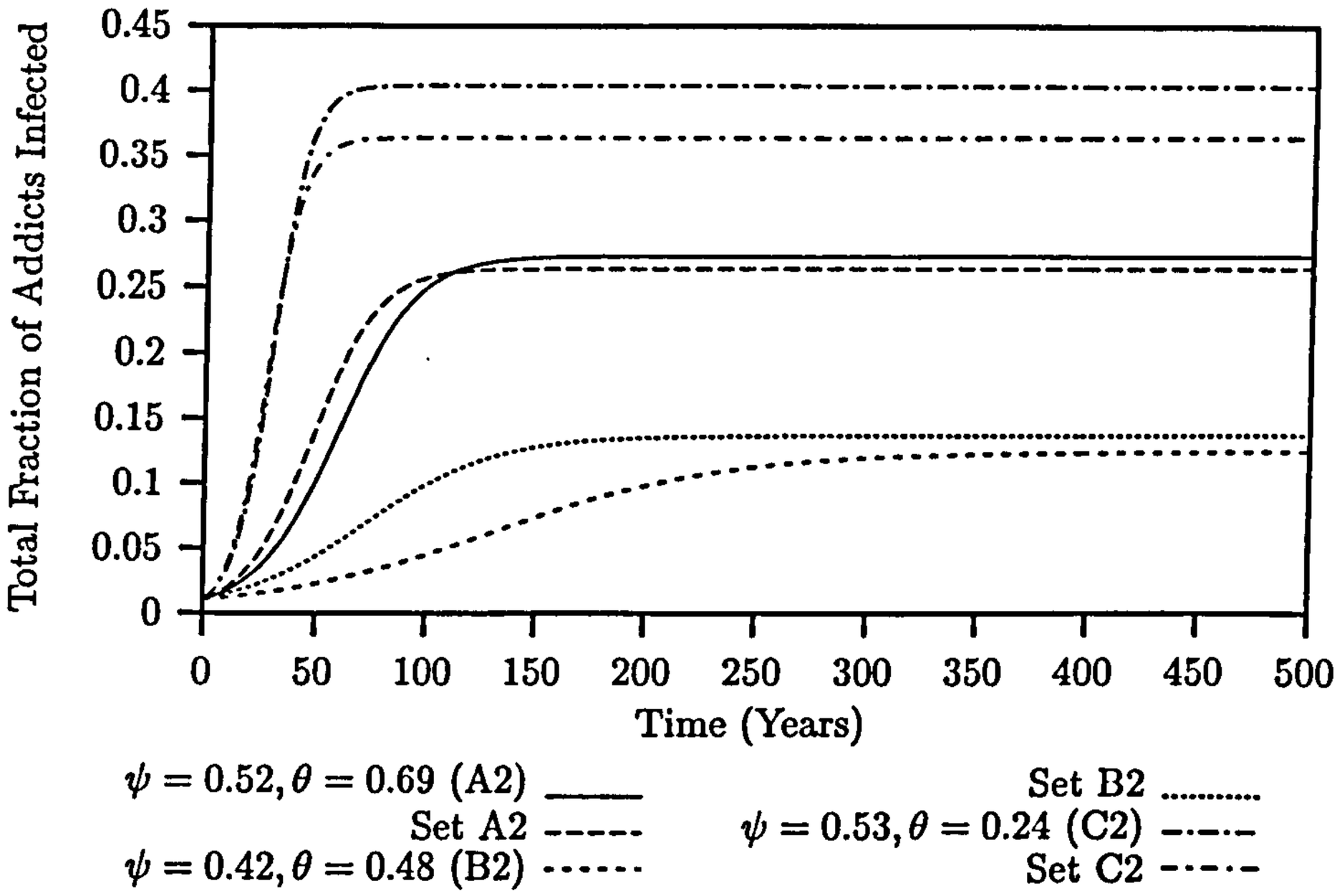
Using eqns (7.5) and (7.6) to equate the viral load in these models reduces our calibration to requiring that

$$\frac{\psi}{\theta} = 1,$$

in the single stage model. This is what we would expect since the Optimistic Model is an extension of the case where $\psi = \theta = 1$. Again these consistent results suggest that this calibration method is sensible, however note that equating R_0^K and R_0^{GEN} does not uniquely identify individual values of ψ and θ but rather only the value of ψ/θ. However this is not a problem as we can determine a unique value for ψ as in the previous calibration method by using eqn (7.45). For example to use this calibration method for any given set of interaction assumptions we first estimate the parameters in each of our models with the exception of ψ and θ in the single stage model, and α_1, α_2 and α_3 in the General Model. We then use eqns (7.5) and (7.6) to find the appropriate value of α_2 (which also gives α_1 and α_3) in the General Model. We then use eqn (7.45) to find the value of ψ corresponding to the set of $p^{\star}_{0jk}$ terms for $j,k = 1,2,3$ in the General Model and use this together with $R_0^K = R_0^{GEN}$ to find the corresponding value of θ in the single stage model.

We now illustrate our current calibration method using the $p^{\star}_{ijk}$ values in Table 7.1. As before we require that $\alpha_2 = 0.00272$, $\alpha_1 = 0.0136$ and $\alpha_3 = 0.00816$ and sets

Figure 7.10: Single Stage v Three Stage (Fully General) : Calibration Method Two



A2-C2 correspond to ψ values of 0.52, 0.42 and 0.53 respectively (using eqn (7.45) as previously). Equating R_0^K and R_0^{GEN} we find that sets A2-C2 correspond to θ values of $0.69, 0.48$ and 0.24 respectively. We therefore have that sets A2-C2 correspond to $(0.52, 0.69), (0.42, 0.48)$ and $(0.53, 0.24)$ for (ψ, θ) respectively. These (ψ, θ) pairs are similar but not identical to those used earlier. We now illustrate simulations of our calibrated models in a similar fashion to Figure 7.9 in the previous section. Figure 7.10 shows a similar pattern of behaviour to that in Figure 7.9 with the main difference being that rather than the three stage model using set A2 having a slightly higher long term prevalence level than its single stage equivalent, it now has a slightly lower long term prevalence level. However the main observation from these simulations is again the lack of a simple summary to describe the comparison between our three stage and single stage models. As previously we find that whether or not three stage infectivity increases the prevalence of disease depends on the individual addict-needle interaction assumptions in our models.

7.9 Addict-Needle Interactions: A Final Remark

Our main purpose for comparing the prevalence of HIV suggested by our various three stage models with existing single stage models has been to establish whether or not a

three stage infectious period increases the level of disease among addicts. Unfortunately the previous sections have shown that moving to a three stage infectious period alone does not imply that an increase or decrease in prevalence will occur, but rather shows that assumptions relating to how addicts and needles interact must also be taken into account. From a practical point of view this is less than ideal since of all the parameters in our single stage and three stage models, those concerned with addict-needle interactions are the most difficult to accurately estimate. This is a particular problem in the General Model where we have so many $p^{\star}_{ijk}$ terms.

We have previously argued that a number of intuitive inequalities should exist between the $p^{\star}_{ijk}$ terms in the General Model. The more constraints and conditions we have between the $p^{\star}_{ijk}$ terms the more we can narrow down the parameter space. By considering the single stage model in eqns (7.43)-(7.44) and the calibration method discussed in Section 7.8.2 we can argue for an additional condition between a number of the $p^{\star}_{ijk}$ terms. Consider the terms θ and ψ. We have denoted by θ the probability that an infectious needle adopts the infectivity characteristics of the current uninfectious user. We have denoted by ψ the probability that an uninfectious needle adopts the infectivity characteristics of the current infectious user. These two probabilities refer to biologically similar events, namely that a needle adopts the infectivity of the current user. Since a needle can only be either infectious or uninfectious it then follows that in order to treat interactions between uninfectious needles and infectious addicts and infectious needles and uninfectious addicts consistently we should have that $\theta = \psi$.

We can use the assumption that $\theta = \psi$ to construct a constraint which allows us to choose consistent values for some of the $p^{\star}_{ijk}$ terms in the General Model. Using $R_0^K = R_0^{GEN}$ we have that

$$\begin{aligned}
\frac{\lambda\alpha}{\mu+\delta} &= \frac{\lambda p^{\star}_{011}}{(\mu+\delta_1)}\left\{\frac{\alpha_1}{1-p^{\star}_{101}} + \frac{p^{\star}_{103}}{1-p^{\star}_{101}}\left[\frac{\alpha_3}{1-p^{\star}_{303}} + \frac{p^{\star}_{302}}{(1-p^{\star}_{303})}\frac{\alpha_2}{(1-p^{\star}_{202})}\right]\right. \\
&\quad \left. + \frac{p^{\star}_{102}}{(1-p^{\star}_{101})}\frac{\alpha_2}{(1-p^{\star}_{202})}\right\} \\
&\quad + \frac{\lambda}{(\mu+\delta_1)}\left[p^{\star}_{012} + \frac{\delta_1 p^{\star}_{022}}{(\mu+\delta_2)} + \frac{\delta_1\delta_2 p^{\star}_{032}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\frac{\alpha_2}{(1-p^{\star}_{202})} \\
&\quad + \frac{\lambda}{(\mu+\delta_1)}\left[p^{\star}_{013} + \frac{\delta_1\delta_2 p^{\star}_{033}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\left[\frac{\alpha_3}{1-p^{\star}_{303}} + \frac{p^{\star}_{302}}{(1-p^{\star}_{303})}\frac{\alpha_2}{(1-p^{\star}_{202})}\right].
\end{aligned} \tag{7.49}$$

We have already shown above that the Optimistic Model satisfies eqn (7.49). This is sensible since by construction this model treats interactions between infectious addicts and uninfectious needles, and uninfectious addicts and infectious needles in the same fashion. Since we expect eqns (7.5) and (7.6) also to be satisfied we require that

$$\left[\frac{\alpha_1}{\mu+\delta_1}+\frac{\alpha_2\delta_1}{(\mu+\delta_1)(\mu+\delta_2)}+\frac{\alpha_3\delta_1\delta_2}{(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\right]=$$

$$\begin{aligned}
&\frac{p^\star_{011}}{(\mu+\delta_1)}\left\{\frac{\alpha_1}{1-p^\star_{101}}+\frac{p^\star_{103}}{1-p^\star_{101}}\left[\frac{\alpha_3}{1-p^\star_{303}}+\frac{p^\star_{302}}{(1-p^\star_{303})}\frac{\alpha_2}{(1-p^\star_{202})}\right]\right.\\
&\left.+\quad\frac{p^\star_{102}}{(1-p^\star_{101})}\frac{\alpha_2}{(1-p^\star_{202})}\right\}\\
&+\quad\frac{1}{(\mu+\delta_1)}\left[p^\star_{012}+\frac{\delta_1 p^\star_{022}}{(\mu+\delta_2)}+\frac{\delta_1\delta_2 p^\star_{032}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\frac{\alpha_2}{(1-p^\star_{202})}\\
&+\quad\frac{1}{(\mu+\delta_1)}\left[p^\star_{013}+\frac{\delta_1\delta_2 p^\star_{033}}{(\mu+\delta_2)(\mu+\delta_3)}\right]\left[\frac{\alpha_3}{1-p^\star_{303}}+\frac{p^\star_{302}}{(1-p^\star_{303})}\frac{\alpha_2}{(1-p^\star_{202})}\right],
\end{aligned}$$

(7.50)

where α_2 (and therefore α_1 and α_3) are chosen to satisfy the calibration eqns (7.5) and (7.6). Hence we can use eqn (7.50) as a guide to choosing $p^\star_{ijk}$ values such that interactions between uninfectious addicts and infectious needles, and infectious addicts and uninfectious needles are treated consistently in the General Model.

Hence to summarise, if we believe that $\theta = \psi$ is an appropriate assumption in single stage infectivity models then we can use this fact together with the various calibration equations shown above to further narrow down the parameter space of our General Model. This concludes our comparisons between the long term prevalence of disease in our three stage and single stage models. We now move on to briefly examining the effect that different control strategies have on the spread of HIV.

7.10 Three Stage Infectivity and Control Strategies

We now investigate the effect of control strategies such as needle exchange programs and improved needle cleaning when we move from single stage infectivity to three stage infectivity. We are interested in both the size and nature of the effect caused by these interventions and also whether they are as effective as suggested by the original Kaplan and O'Keefe Model. However before we do this we first make a brief comment as to

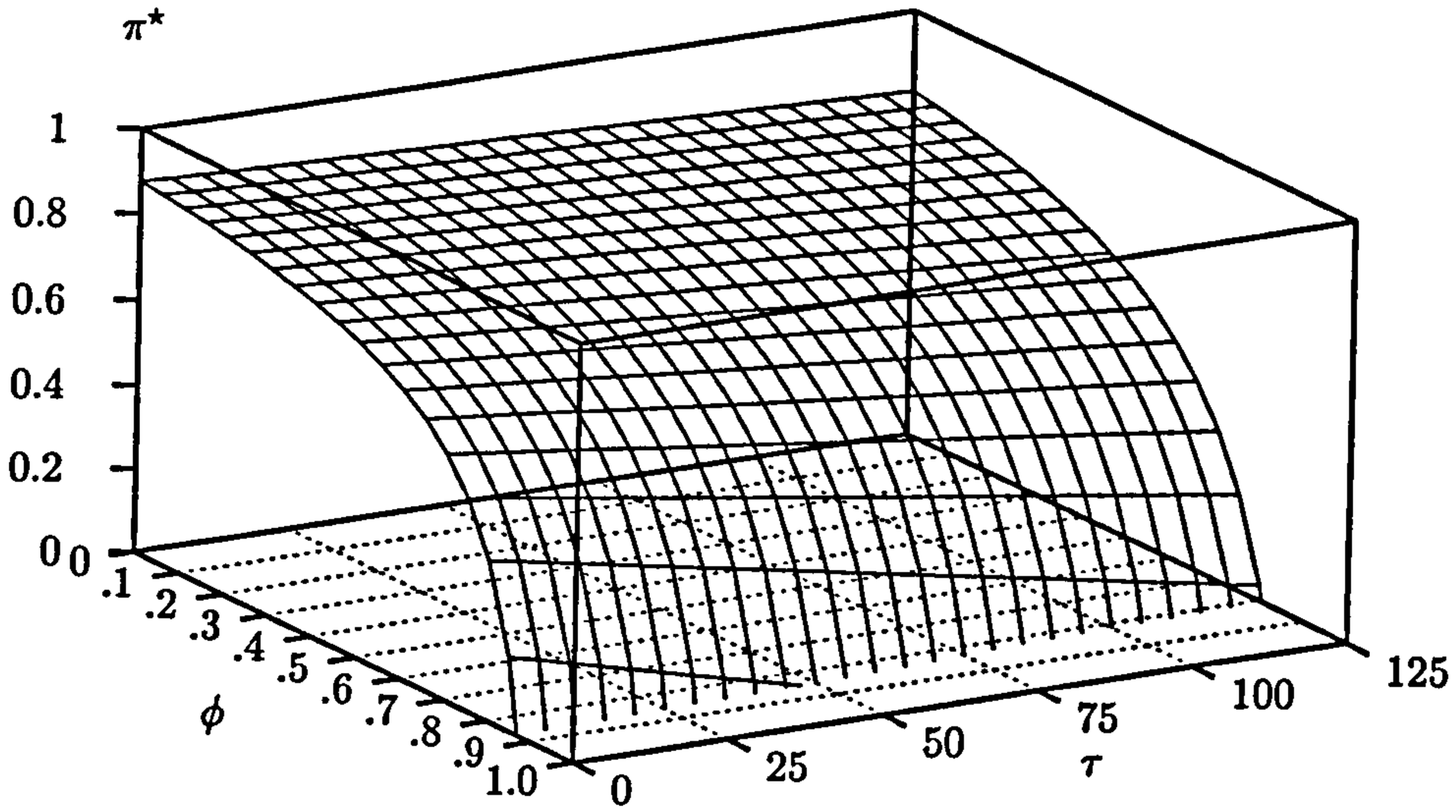
the impact of the gallery ratio, that is the ratio of addicts to needles, on our various models.

Intuitively we might expect that the gallery ratio, denoted in our models by γ, would have a significant impact on the long term level of disease. For example if γ is large then each individual needle is used (or shared) among more addicts and therefore we might expect that this would increase the spread of disease. Kaplan (1989a) examines the effect of γ in a simple homogeneous model of needle sharing and finds that while this parameter does have an impact on the speed at which the disease spreads it has no effect on either R_0 or the endemic equilibrium level of disease (this parameter cancels out in these expressions). We do not illustrate simulations of our three stage models for different values of γ, however we also found that while γ does not cancel out in the expressions of either R_0 or the endemic equilibrium solutions in our models it has only a very minor quantitative effect (at least over the range of values which we examined).

In the previous chapters we have demonstrated two main aspects of our models. Firstly there exists a critical threshold parameter which determines whether or not an epidemic takes off, and secondly when an epidemic does occur the prevalence of disease tends to an endemic steady state. Hence it seems natural to use the critical threshold parameter, R_0, and the endemic equilibrium solution in our various models to assess the impact on the spread of disease caused by introducing a formal needle exchange program and improved needle cleansing practices.

We first examine the qualitative impact on the long term prevalence of disease in our models caused by improved needle cleaning and the implementation of a needle exchange program. Figure 7.11 illustrates simultaneously the effect of cleaning and needle exchange on the behaviour of the total long term prevalence of disease among addicts in the Simple Model (denoted by $\pi^\star$). This is of practical interest as needle exchange programs do not only allow needles to be exchanged but also offer guidance and advice on risk reduction practices (such as reducing the sharing of needles and cleaning needles prior to use). Therefore it may be that by participating in such a program addicts increase both the rate at which needles are exchanged and the frequency with which they clean needles prior to injection. In Figure 7.11 we have assumed that addicts clean needles prior to use with a probability of between 0 and 1.0, (we have denoted this probability by ϕ), and needles are exchanged between 0 and 125 times a year on average (we have denoted the needle exchange rate by τ). All other model parameters are as in Table 2.1. From the figure the relationship between $\pi^\star$ and

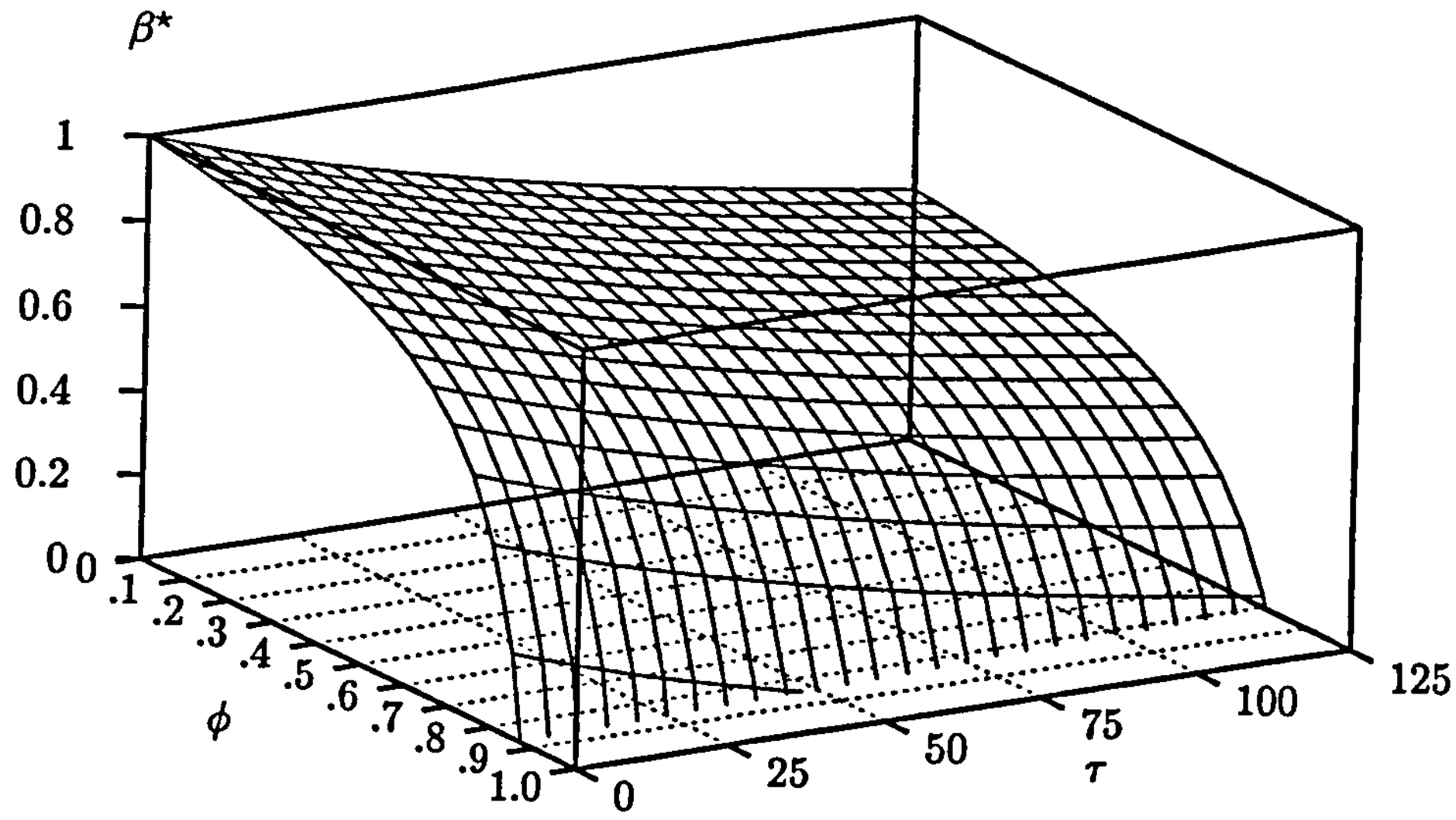
Figure 7.11: Impact of ϕ and τ on the Long Term Prevalence in Addicts (Simple Model)



ϕ appears to be non-linear with $\pi^\star$ much more sensitive to changes in ϕ for ϕ large, whereas the relationship between $\pi^\star$ and τ appears to be linear (as expected from eqn (2.8) in Chapter 2). In both cases we have that increasing the probability of cleaning or the needle exchange rate decreases $\pi^\star$ as intuitively must be the case (again this also follows from eqn (2.8)).

Figure 7.12 is similar to Figure 7.11 but shows the behaviour of the total prevalence of disease among needles in the Simple Model for the previous ranges of ϕ and τ. This figure suggests that $\beta^\star$ is much more sensitive to changes in ϕ and τ when both of these are small, in other words when addicts do not clean needles or needles are never exchanged. Apart from this interaction effect it appears that the relationship between cleaning and $\beta^\star$ is broadly similar to that with $\pi^\star$. However this figure suggests that increasing τ reduces the value of $\beta^\star$ more than it reduces the value of $\pi^\star$ (all other things being equal). For example if we fix $\phi = 0.64$ and $\tau = 15.53$ per year we find that $\pi^\star = 0.633$ and $\beta^\star = 0.675$. If we now increase τ to 121.7 per year (which corresponds to a mean needle circulation time of 3 days, as estimated by Kaplan, 1995) we find that now $\pi^\star = 0.471$ and $\beta^\star = 0.348$, a decrease of 26% and 49% respectively. Hence needle exchange has a much greater impact on the level of disease among needles than among addicts. Simulations of our other three stage models using a variety of different parameter estimates suggest that this is the case in general. Kaplan (1995) demonstrated that introducing a needle exchange scheme could cause a decrease of up

Figure 7.12: Impact of ϕ and τ on the Long Term Prevalence in Needles (Simple Model)



to 33% in the prevalence of disease among needles. Kaplan then implied that this would correspond to a similar decrease of 33% in the prevalence of disease among addicts. Our models suggest that may not be the case and that the corresponding decrease in the prevalence of disease among addicts may be much more modest.

Figures 7.13-7.15 show the relationship between the total long term prevalence of disease in addicts and needles for the parameters τ and ϕ using the Pessimistic, Optimistic and General Models (where the latter uses $p^{\star}_{ijk}$ set $C1$). We assume that $\tau = 15.53$ per year when varying ϕ and that $\phi = 0.64$ when varying τ, all other model parameters are as in Table 3.1. These figures show a number of interesting features. Firstly as was the case in the Simple Model it appears that needle exchange is approximately linearly related to the long term prevalence of disease in both addicts and needles. In fact by examining the expressions for the equilibrium prevalence of disease amongst addicts and needles in the Optimistic Model it is straightforward to show that these are linearly related to the needle exchange rate τ. Similarly we find that the probability of needle cleaning is related non-linearly to the long term prevalence of disease in both addicts and needles with changes in ϕ for ϕ close to one having relatively more impact.

It is interesting to note that the relationship between τ and the long term prevalence of disease in needles and does not necessarily mimic that of the relationship between τ and the long term prevalence of disease in addicts. Moreover it appears that (at least

Figure 7.13: $(\pi^\star, \beta^\star)$ in Pessimistic Model: $\pi^\star$ = Solid, $\beta^\star$ = Dashed

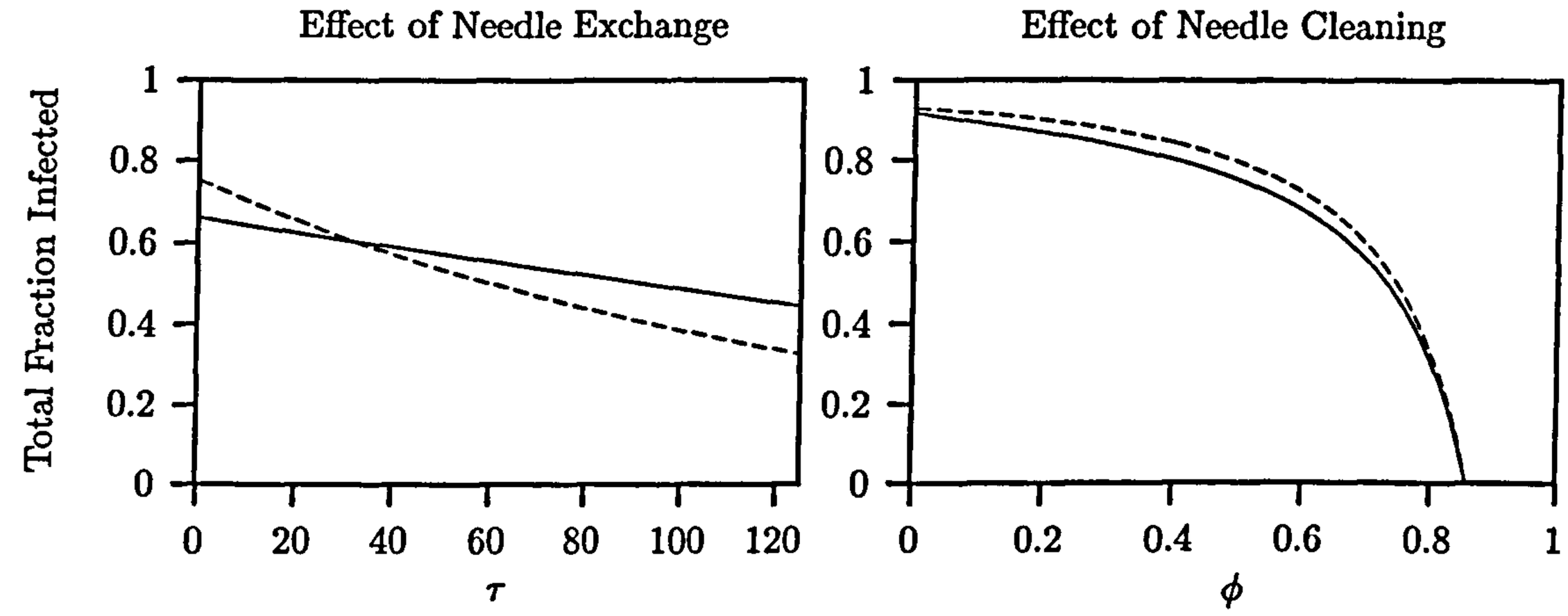


Figure 7.14: $(\pi^\star, \beta^\star)$ in Optimistic Model: $\pi^\star$ = Solid, $\beta^\star$ = Dashed

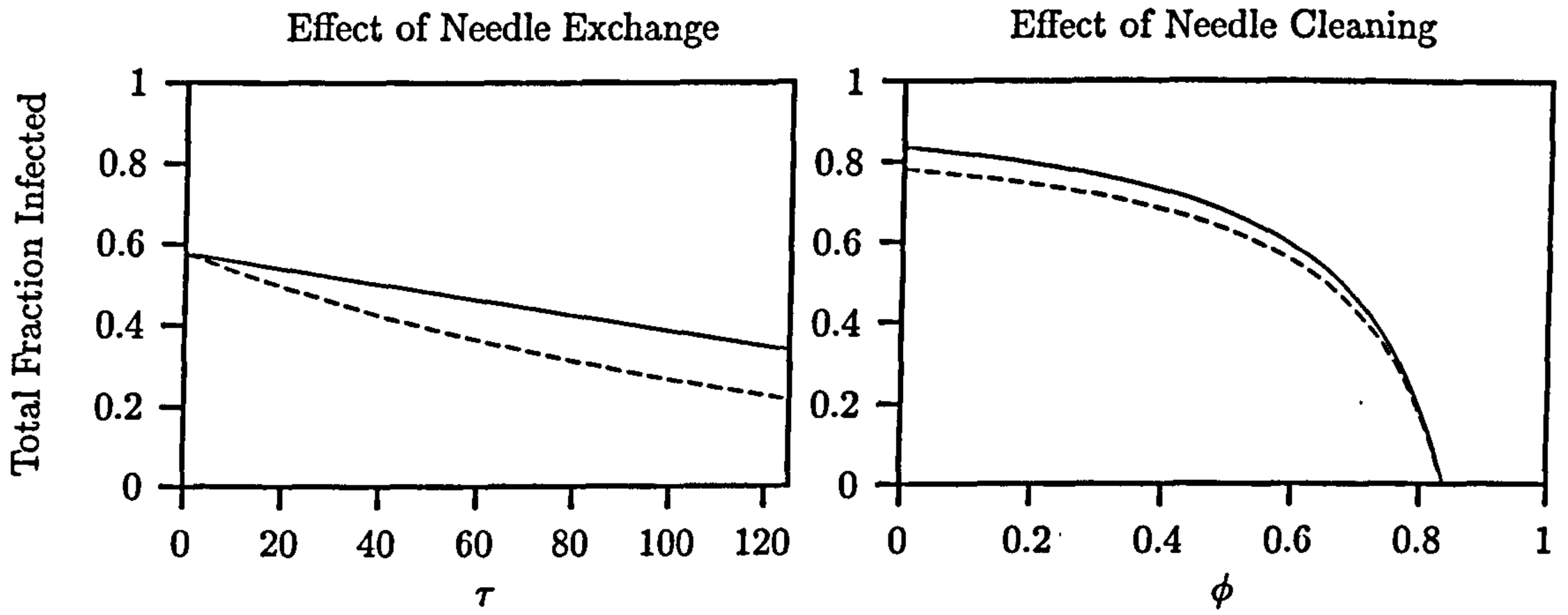
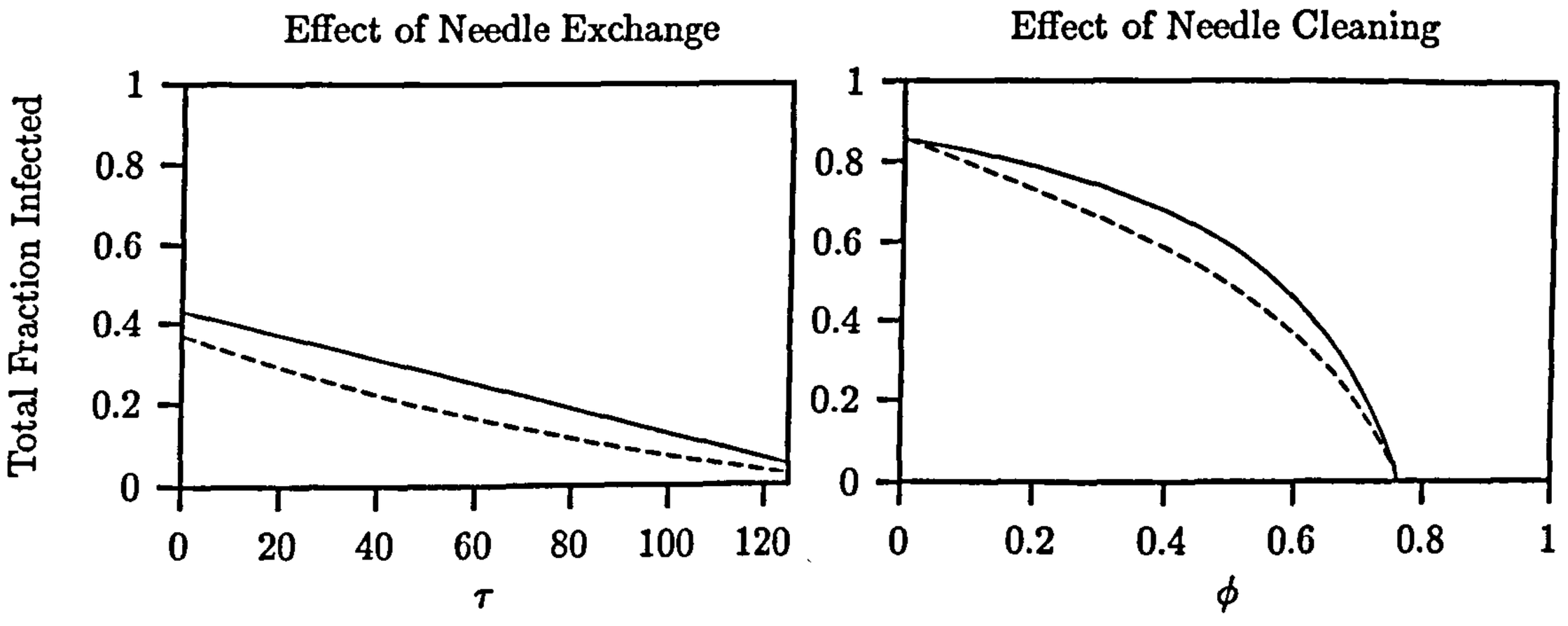


Figure 7.15: $(\pi^\star, \beta^\star)$ in General Model (using C1): $\pi^\star$ = Solid, $\beta^\star$ = Dashed



in the Pessimistic and Optimistic Models) the relative decrease in $\beta^\star$ due increasing the rate of needle exchange is larger than that observed in $\pi^\star$. This is consistent with our previous discussion regarding the Simple Model, therefore we have again that needle exchange is more effective in reducing the level of disease among needles than that among addicts.

To get an idea as to the level of quantitative decrease we might expect given the introduction of the previous control strategies into a variable infectivity needle sharing environment we can use the Restricted General Model together with the five sets of $p^\star_{ijk}$ terms in Table 5.4. Table 7.2 shows the relative decrease in the long term prevalence of disease in addicts caused by increasing the needle exchange rate. The percentage

Table 7.2: Relative Reduction in $\pi^\star$ due to Needle Exchange

$p^\star_{ijk}$ Selection	Mean Needle Circulation Time			
	23.5 Days	5 Days	4 Days	3 Days
Optim	0.0%	20.0%	26.4%	37.0%
A	0.0%	18.9%	25.0%	35.0%
B	0.0%	17.8%	23.4%	32.8%
C	0.0%	16.6%	21.8%	30.6%
Pessim	0.0%	15.8%	20.7%	29.1%

reduction in Table 7.2 uses a mean needle circulation time of 23.5 days as the baseline, this corresponds to a natural needle turnover rate of 15.53 per year as estimated by Kaplan (1995). As already mentioned Kaplan estimates that once a formal needle exchange scheme has been established the mean needle circulation time can drop to between 3 and 5 days. Table 7.2 shows that as the addict-needle interaction assumptions move from the Optimistic Model to the Pessimistic Model the relative effectiveness of needle exchange decreases. Examining the relative decrease in $\pi^\star$ in Table 7.2 for sets A-C it seems reasonable to conclude that a needle exchange scheme, which reduces the mean needle circulation time to three days (on average), will reduce the long term prevalence of disease in addicts by between 30%-35%. This is very similar to the decrease claimed by Kaplan and O'Keefe (1993) and Kaplan (1995) using single stage infectivity models. This is further good evidence that as a control strategy to prevent the spread of HIV, needle exchange programs are highly beneficial.

We now examine the effect of improved cleaning on the long term prevalence of HIV in addicts. Table 7.3 is similar to Table 7.2 but instead of increasing τ in the Restricted General Model we now increase ϕ. The percentage reduction in Table 7.3 uses $\phi = 0.64$

Table 7.3: Relative Reduction in $\pi^\star$ due to Cleaning

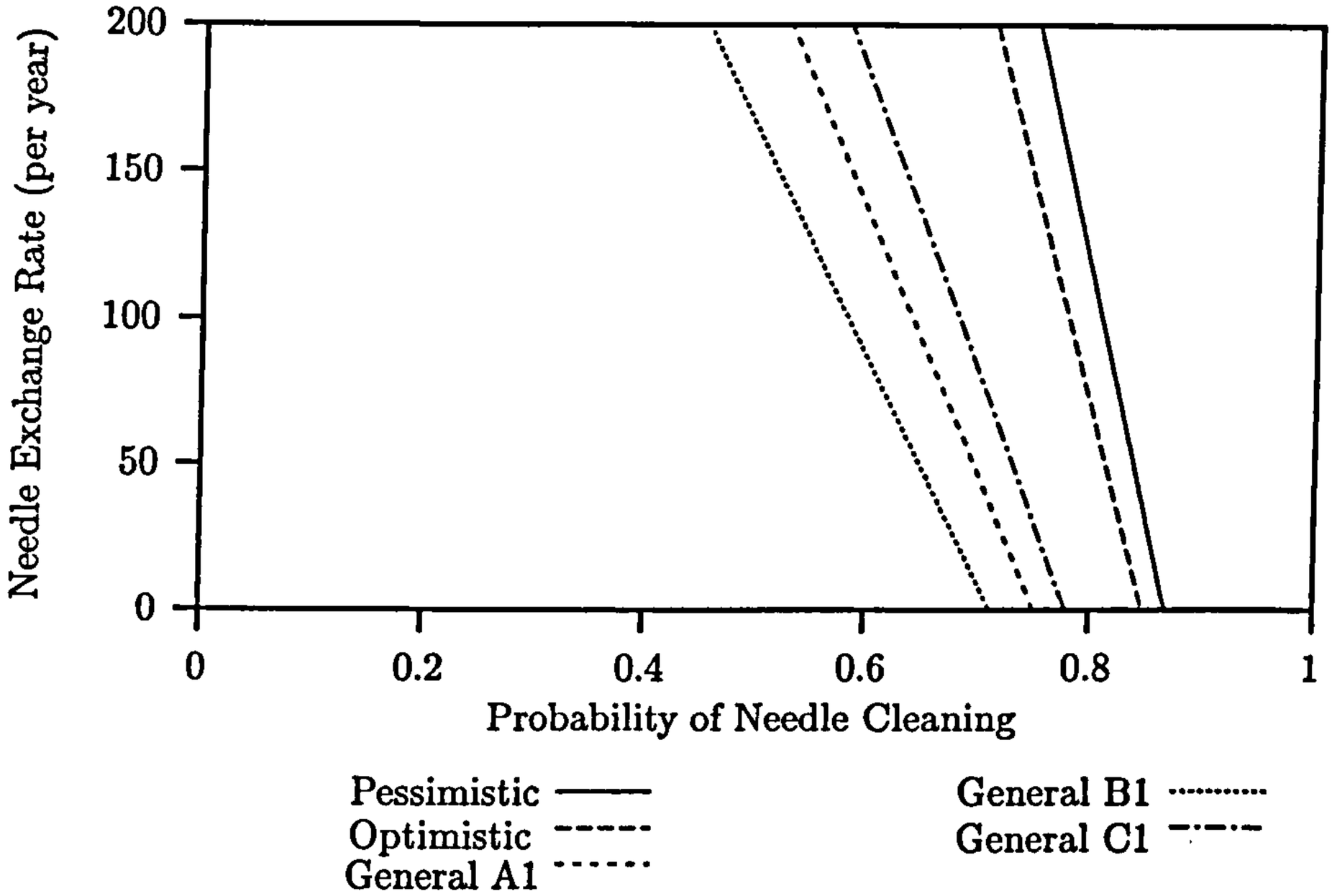
	Probability Needle Cleaned prior to use			
p_{ijk} Selection	0.64	0.707	0.773	0.84
Optim	0.0%	18.9%	48.9%	100.0%
A	0.0%	17.9%	46.4%	98.6%
B	0.0%	16.8%	43.4%	92.3%
C	0.0%	15.7%	40.5%	86.0%
Pessim	0.0%	14.8%	38.4%	82.6%

as the baseline, in other words we assume that normally 64% of addicts successfully clean needles prior to use. It is difficult to estimate the actual level of cleaning which occurs in practice so the values in Table 7.3 are just a guide to show the size of effect which improved cleaning can produce. The main point of interest in the table is the very large decrease in $\pi^\star$ caused by a relatively small increase in ϕ. Intuitively the reason for this is that cleaning is the most direct form of intervention to reduce the spread of HIV via needle sharing. For example, if some uninfected addicts always successfully clean needles prior to use, then they themselves will remain uninfectious and moreover they increase the amount of uncontaminated needles in circulation (since these addicts will always leave needles in an uncontaminated state).

We have briefly examined the qualitative and quantitative relationship between the long term prevalence of disease and the introduction of improved needle cleaning and needle exchange. We now look at the relationship between these control measures and R_0. This quantity is fundamental in determining whether or not an epidemic occurs and moreover when R_0 is relatively small it also has a basic influence on the long term level of disease in the population. For example in the Optimistic Model we have explicitly that $\pi^\star = 1 - (1/R_0)$. Even though this is one of the simplest of our three stage infectivity models numerical studies suggest that very similar relationships hold between R_0 and $\pi^\star$ for all our other models. In particular it appears to be the case that when R_0 is close to the critical threshold of unity $\pi^\star$ is very sensitive to changes

Figure 7.16: Critical Threshold of $R_0(\phi, \tau) = 1$ for Varying λ and p^*_{ijk}

$\lambda = 246.22$ per year



in this parameter.

Using eqns (3.7) and (4.7) respectively and writing $R_0 = R_0(\phi, \tau)$ it is straightforward to show that to eradicate the disease using needle exchange if the probability of needle cleaning ϕ is fixed we require that

$$\tau \geq \tau_0 = \lambda\gamma\hat{\theta}(R_0(\phi, 0) - 1),$$

where $\hat{\theta} = 1$ in the Optimistic Model and $\hat{\theta} = \phi$ in the Pessimistic Model. Similarly to eradicate the disease in the Optimistic and Pessimistic Models using cleaning if the needle exchange rate τ is fixed we require that

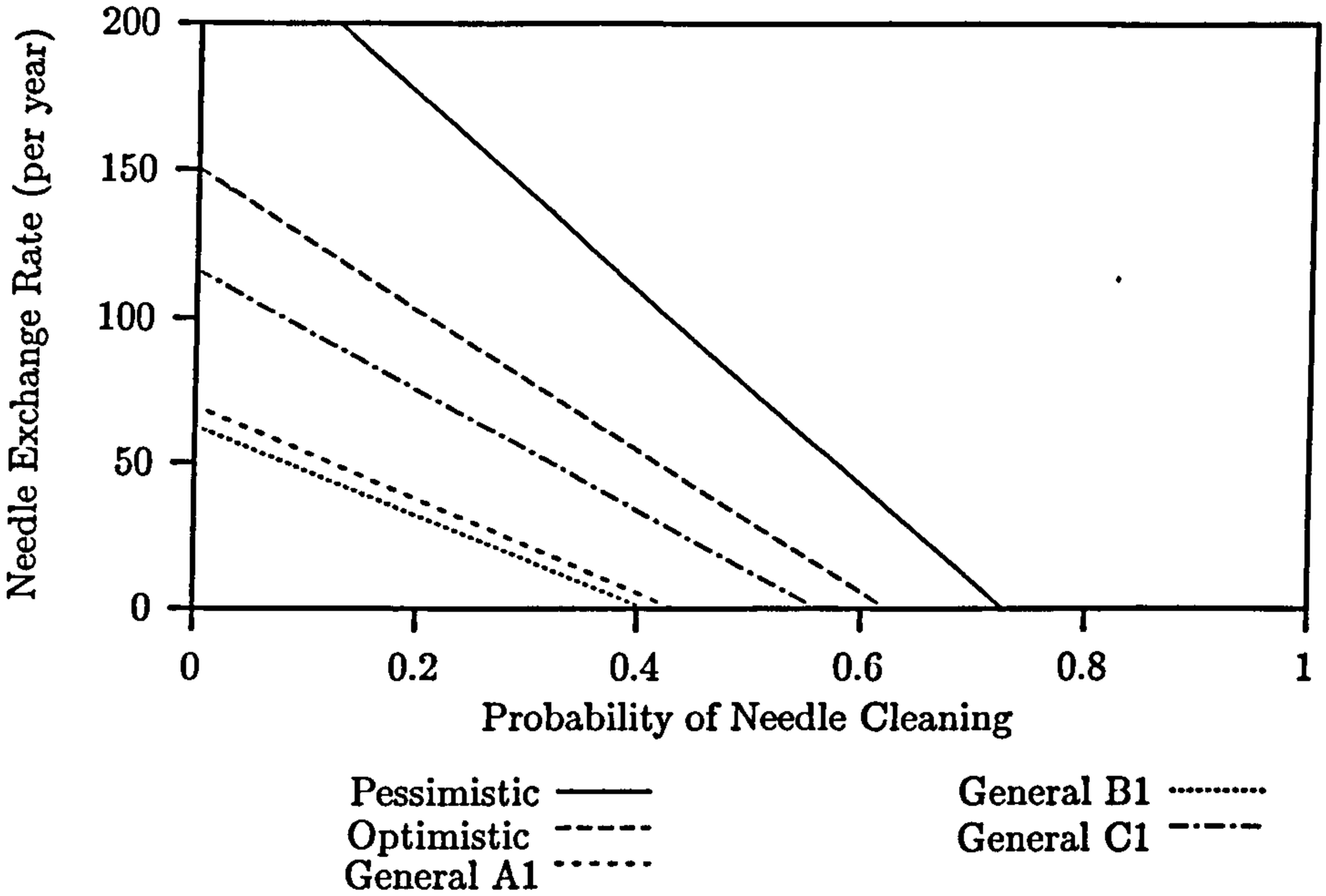
$$\phi \geq \phi_0 = \frac{R_0(0, \tau) - 1}{R_0(0, \tau) + \frac{1-\theta}{\tau+\theta}},$$

where again $\hat{\theta} = 1$ in the Optimistic Model and $\hat{\theta} = \phi$ in the Pessimistic Model. In principal we could use the expression for R_0 in the General Model to show similar results for the Restricted General and General Models, however separating out the τ and ϕ terms in this expression is very much more difficult and we did not attempt it.

Figure 7.16 shows the degree of needle exchange and needle cleaning required to reduce R_0 in the Optimistic, Pessimistic and selected cases of the General Model down to the critical threshold of unity. Apart from ϕ and τ all other parameters are fixed

Figure 7.17: Critical Threshold of $R_0(\phi, \tau) = 1$ for Varying λ and p^*_{ijk}

$\lambda = 100.0$ per year



at the values in Table 3.1. Each curve divides the (ϕ, τ) plane into two with disease persistence on the side with the lower values of (ϕ, τ) and disease eradication on the other side. As we might have expected from Figure 5.5 in Chapter 5 more intervention is required to reduce R_0 down to unity in the Pessimistic Model followed by the Optimistic Model and then the General Model with C1, A1 then B1. The figure suggests that if $\phi = 0.64$ as previously estimated then introducing a needle exchange scheme which increases τ to approximately 121.7 per year on average will eradicate disease in the General Model with B1 and A1 but none of the other models shown. Figure 7.17 is similar to Figure 7.16 but now the rate at which addicts inject, λ, has been decreased down to 100 per year. It is now clear that with $\phi = 0.64$ and $\tau = 121.7$ per year then disease can be eradicated in each of the models shown. This is an important point as this illustrates clearly the importance of lowering the rate at which addicts inject irrespective of the current level of needle exchange and cleaning.

7.10.1 Summary of Control Strategies

We have found that the effect of introducing a needle exchange program into our three stage infectivity models is broadly similar to that suggested by the single stage Kaplan and O'Keefe Model. This is not surprising as whether or not a needle is exchanged

should not depend on what stage of infectivity the addict exchanging the needle is currently in, or the infectious state of the needle being exchanged. However we do find that the effectiveness of this control policy in reducing the prevalence of disease among addicts does to some extent depend on assumptions made relating to how addicts and needles of different infectivity levels interact. More importantly we also found that while needle exchange is very effective in reducing the level of disease among needles the resulting reduction in the level of disease among addicts may be much smaller. We demonstrated that improved needle cleaning can substantially decrease the prevalence of disease among addicts, however we have commented that this policy may be relatively more effective when introduced into a population where some degree of cleaning already occurs.

With the parameter estimates used in this thesis it is unlikely that needle exchange and improved needle cleaning alone could eradicate the disease completely in all our models; we require additional measures such as a reduction in the rate at which needles are shared. As previously mentioned when addicts participate in an exchange program they can receive counselling on their risk taking practices which may help in lowering the rate at which addicts share needles. An additional measure which may prove useful would be to run an HIV testing program in conjunction with an existing needle exchange program. The motivation behind such a program is that addicts who test positive for HIV will hopefully substantially reduce the rate at which they share needles. We examine the impact of HIV testing in Chapter 8.

7.11 Summary of Practical Implications

In this chapter we examined what new information can be obtained from comparing our three stage infectivity models with existing single stage infectivity models. We first compared the long term prevalence of HIV in single stage models with that of equivalent three stage models. We noted that it is not straightforward to compare single stage and three stage models in order to isolate the effect of splitting infectious addicts and infectious needles each into three infectious classes. The way addicts and needles interact is a major complication in this process. We compared the Kaplan and O'Keefe Model with the Restricted General Model and found that these models have comparable upper and lower bounds for the long term prevalence of disease. We also compared an extended version of the Kaplan and O'Keefe Model with the General

Model. We found that in general, moving to three stage infectivity can either increase or decrease the long term prevalence of disease, depending on the particular addict-needle interaction assumptions in our models.

We briefly considered the effect of control strategies such as needle exchange programs and improved needle cleaning in three stage infectivity models compared to single stage infectivity models. We found that these control measures were just as effective as in single stage models, in particular the well publicised claim (Kaplan and O'Keefe, 1993, Kaplan, 1994, and Kaplan, 1995) that introducing a formal needle exchange can reduce the long term prevalence of disease by approximately 33% still appears to hold. However we did note that assumptions relating to how addicts and needles of different infectivity levels interact do have an impact on the effectiveness of this control strategy.

This chapter concludes our direct interest in three stage infectivity models. We now move on to another area of interest in the modelling of HIV among intravenous drug users. It has been suggested by Greenhalgh and Hay (1997) that testing addicts for HIV might be a worthwhile control strategy. However in contrast Kretzschmar and Wiessing (1998) claim that HIV testing is of limited use in reducing the spread of HIV. The following chapter investigates the behaviour of models which incorporate HIV testing and examines whether this could be an effective control strategy.

Chapter 8

The Effect of Testing Addicts for the Presence of HIV

8.1 Introduction

In Greenhalgh and Hay (1997) an original model due to Kaplan (1989a) is extended to include a variety of additional features, in particular the testing of addicts for HIV. Greenhalgh and Hay found that if addicts who are aware of their HIV positive status (through participating in an HIV test) reduce the rate at which they share needles, then this can dramatically reduce both the speed at which the disease spreads, and the equilibrium level of disease incidence. While this result has potentially important practical significance, Greenhalgh and Hay acknowledge that the treatment of HIV testing in their model could be improved and suggest a method of doing so.

In this chapter we first discuss a model similar to that examined by Greenhalgh and Hay (1997). We then extend this model to deal with HIV testing in more realistic fashion and study the behaviour of this more complex model using both simulation and analytical results. We later compare the original Greenhalgh and Hay Model with this extended model (which we shall refer to as the HIV Test Model) to ascertain whether our more realistic treatment of HIV testing has any effect on the long term prevalence of disease compared to the original model. We then briefly use our model to examine whether HIV testing could be an effective control strategy against the spread of HIV among injecting drug users. Next we extend our model to incorporate a three stage infectious period using the Optimistic and Pessimistic Models discussed in Chapters 3 and 4 respectively. We conduct a brief study of the behaviour of these models before

investigating the effect that different addict-needle interaction assumptions and relative infectivity assumptions have on the effectiveness of HIV testing. The chapter concludes with a summary of the main points.

8.2 The Greenhalgh and Hay Model

We now outline the approach to HIV testing used by Greenhalgh and Hay (1997). Using the Kaplan and O'Keefe Model illustrated in eqns (1.1)-(1.2) as a starting point, the simplest way of incorporating HIV testing is to split the population of infectious addicts into two groups, those who have tested positive for HIV and those who have not. This is the approach adopted by Greenhalgh and Hay. They assumed that at all times a fixed proportion p of the infected addict population are aware of their infectious status. The remaining proportion $1-p$ of infected addicts are not aware of their infectious status. It is assumed that addicts who are not aware of their infectious status have a shared injection rate of λ_1 per unit time, the same sharing rate as for susceptible addicts, whereas those addicts who have tested positive for HIV have a shared injection rate of λ_2 per unit time. The motivation for HIV testing comes from the fact that we expect λ_1 to be greater than λ_2. It is easy to modify the Kaplan and O'Keefe Model in eqns (1.1)-(1.2) to include this treatment of HIV testing, hence we have that

$$\frac{d\pi}{dt} = (1-\pi)\lambda_1\beta\alpha(1-\phi) - \pi(\mu+\delta), \tag{8.1}$$

and

$$\frac{d\beta}{dt} = (1-\beta)\gamma\Big[(1-p)\lambda_1 + p\lambda_2\Big]\pi - \beta\lambda_1\gamma(1-\pi)(1-(1-\theta)(1-\phi)) - \beta\tau. \tag{8.2}$$

Examining eqns (8.1) and (8.2) suggests that they do not differ substantially in structure from Kaplan and O'Keefe's original model. However as we shall discuss in due course these equations are less easy to deal with mathematically than those in Kaplan and O'Keefe's original model or indeed the systems of equations in the Simple, Optimistic or Pessimistic Models.

Greenhalgh and Hay (1997) examine the behaviour of a model very similar to that defined by eqns (8.1)-(8.2). They show analytically that there exists a unique endemic equilibrium if and only if $R_0 > 1$. They show global stability results for the disease-free equilibrium if $R_0 \leq 1$ and local stability results for the disease-free and endemic equilibria if $R_0 > 1$. In the case where $R_0 > 1$ and $(1-p)\lambda_1 + p\lambda_2 > \lambda_1\hat{\theta}$ they also show

global stability results for the endemic equilibrium. In their model they assume that needles circulate among the population indefinitely, (which implies that $\tau = 0$). In the model defined above we have made the more realistic assumption that needles must be replaced eventually (and therefore $\tau > 0$), this is entirely reasonable since at the very least needles have a limited working lifetime. In addition the model defined above assumes that addicts clean needles prior to use rather than after use as in Greenhalgh and Hay (1997). This again is a more realistic assumption as cleaning a needle after use (rather than prior to use) does not serve to protect the current user but rather cleans the needle for the next user. These two modifications are minor and the analytical results demonstrated by Greenhalgh and Hay can easily be extended to cater for these two small extensions.

Greenhalgh and Hay suggest that a more realistic method of modelling the effect of HIV testing would be to assume that all newly infected addicts enter a class where they are unaware of their HIV status, and are subsequently tested for HIV at a constant rate per unit time, after this they enter a second class of infectious addicts who are aware of their HIV status, and this class has a different shared injection rate from the first. In effect this approach allows the proportion of infectious addicts who know their HIV status to fluctuate as the disease spreads among the population rather than remaining static as in the model in eqns (8.1)-(8.2). Therefore rather than a fixed proportion p of infectious addicts being aware of their HIV positive status we now have that at time t a proportion $p(t)$ of infectious addicts are aware that they are infected with HIV. This is a more realistic method of dealing with HIV testing and it is this which we now examine.

8.3 Model Derivation

We now derive the differential equations which define the spread of HIV among an intravenous drug addict population where addicts are tested for HIV at rate δ_t per unit time and a positive test results in a change of the needle sharing rate from λ_1 per unit time to λ_2 per unit time. HIV positive addicts develop AIDS at per capita rate δ per unit time and on developing AIDS they immediately leave the sharing, injecting population. We derive three equations: one for infectious addicts who have not yet tested positive for HIV; one for infectious addicts who have tested positive for HIV; and one for the infectious needles used by both types of addicts. We shall refer to infectious

addicts who have not yet tested positive for HIV as type one addicts and those who have tested positive as type two addicts.

The number of type one infected addicts at time $t + \Delta t$

$$
\begin{aligned}
= \; & \{\text{number of type one infected addicts at time } t\} \\
& + \{(\text{number of uninfected addicts at time } t) \\
& \times (\text{fraction of addicts who inject in } [t, t+\Delta t) \text{ with an infectious needle which is not cleaned prior to use and where transmission of HIV occurs in a single injection})\} \\
& - \{\text{number of type one infected addicts who test positive for HIV, develop AIDS or leave the sharing, injecting population for other reasons in } [t, t+\Delta t)\}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
n\pi_1(t+\Delta t) = \; & n\pi_1(t) + n(1-\pi_1(t)-\pi_2(t))\lambda_1 \Delta t \beta(t)\alpha(1-\phi) - n\pi_1(t)\Delta t(\mu + \delta_t + \delta) \\
& + o(\Delta t).
\end{aligned}
$$

Subtracting $n\pi_1(t)$ from both sides, dividing by $n\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\pi_1}{dt} = (1-\pi_1-\pi_2)\lambda_1\beta\alpha(1-\phi) - \pi_1(\mu+\delta_t+\delta).$$

The number of type two infected addicts at time $t + \Delta t$

$$
\begin{aligned}
= \; & \{\text{number of type two infected addicts at time } t\} \\
& + \{\text{number of type one infected addicts who test positive for HIV in } [t, t+\Delta t)\} \\
& - \{\text{number of type two infected addicts who develop AIDS or leave the sharing, injecting population for other reasons in } [t, t+\Delta t)\}.
\end{aligned}
$$

Thus

$$n\pi_2(t+\Delta t) = n\pi_2(t) + n\pi_1(t)\delta_t\Delta t - n\pi_2(t)(\mu+\delta)\Delta t + o(\Delta t).$$

Subtracting $n\pi_2(t)$ from both sides, dividing by $n\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\pi_2}{dt} = \delta_t \pi_1 - (\mu + \delta)\pi_2.$$

The number of infected needles at time $t + \Delta t$

$$\begin{aligned} = \ & \{\text{number of infected needles at time } t\} \\ & + \{(\text{number of uninfected needles at time } t) \times (\text{fraction of} \\ & \text{needles used by infected addicts in } [t, t+\Delta t))\} \\ & - \{(\text{number of infected needles at time } t) \times (\text{fraction} \\ & \text{of infectious needles used by uninfected addicts in } [t, t+\Delta t) \\ & \text{and left in an uninfected state})\} \\ & - \{\text{number of infected needles exchanged in } [t, t+\Delta t)\}. \end{aligned}$$

Thus

$$\begin{aligned} m\beta(t+\Delta t) = \ & m\beta(t) + m(1-\beta(t))\Delta t\gamma(\pi_1(t)\lambda_1 + \pi_2(t)\lambda_2) \\ & - m\beta(t)\lambda_1 \Delta t\gamma(1-\pi_1(t)-\pi_2(t))(1-(1-\phi)(1-\theta)) \\ & - m\beta(t)\tau\Delta t + o(\Delta t). \end{aligned}$$

Subtracting $m\beta(t)$ from both sides, dividing by $m\Delta t$ and letting $\Delta t \to 0$ we deduce that

$$\frac{d\beta}{dt} = (1-\beta)\gamma(\pi_1\lambda_1 + \pi_2\lambda_2) - \beta\lambda_1\gamma(1-\pi_1-\pi_2)(1-(1-\theta)(1-\phi)) - \beta\tau.$$

Hence the system of differential equations which describes the spread of the disease is:

$$\frac{d\pi_1}{dt} = (1-\pi_1-\pi_2)\lambda_1\beta\alpha(1-\phi) - (\mu+\delta_t+\delta)\pi_1, \tag{8.3}$$

$$\frac{d\pi_2}{dt} = \delta_t\pi_1 - (\mu+\delta)\pi_2, \tag{8.4}$$

$$\text{and} \quad \frac{d\beta}{dt} = (1-\beta)\gamma(\pi_1\lambda_1 + \pi_2\lambda_2) - \beta\lambda_1\gamma(1-\pi_1-\pi_2)(1-(1-\theta)(1-\phi)) - \beta\tau, \tag{8.5}$$

with suitable initial conditions: $0 \leq \pi_1(0), \pi_2(0), \beta(0)$ and $\pi_1(0) + \pi_2(0), \beta(0) \leq 1$.

8.4 Equilibrium and Stability Results

We wish to use the model defined by eqns (8.3)-(8.5) to examine the impact of testing addicts for the presence of HIV. A natural performance measure of the effect of HIV testing is the long term (equilibrium) prevalence of disease predicted by our model. Additionally we are interested in whether our model possess a critical threshold parameter which will predict whether or not the disease will take off if initially present. If so this parameter will give an indication of whether HIV testing alone could produce an effect big enough to eventually eradicate disease among the population. First we define

$$R_0 = \frac{\lambda_1 \alpha(1-\phi)}{(\mu+\delta_t+\delta)(\hat{\tau}_1+\lambda_1\hat{\theta})}\left[\lambda_1 + \frac{\lambda_2\delta_t}{(\mu+\delta)}\right], \tag{8.6}$$

where $\hat{\tau}_1 = \tau/\gamma$ and $\hat{\theta} = 1-(1-\phi)(1-\theta)$. We now investigate the behaviour of our model and our results are summarised in the following theorem:

Theorem 8.1 *There is always the unique equilibrium where there is no disease present in either addicts or needles. If $R_0 \leq 1$ then irrespective of the state of the current epidemic the disease will eventually die out among addicts and needles. If $R_0 > 1$ and if at least one addict or one needle is initially infected then disease will remain persistent among the population indefinitely, moreover there now also exists a unique endemic equilibrium solution which is locally stable.*

Proof.

It is obvious that the disease-free equilibrium is always a solution to eqns (8.3)-(8.5). We now show that an endemic solution exists if and only if $R_0 > 1$. Note that we assume that all model parameters with the exception of ϕ are strictly positive and that $\phi < 1$. From eqns (8.3)-(8.5) we have that

$$(1-\pi^\star)\lambda_1\alpha(1-\phi)\beta^\star = \frac{\pi^\star}{L}(\mu+\delta_t+\delta), \tag{8.7}$$

and

$$(1-\beta^\star)\gamma(\pi_1^\star\lambda_1+\pi_2^\star\lambda_2) = \beta^\star\lambda_1\gamma(1-\pi^\star)\hat{\theta}+\beta^\star\tau, \tag{8.8}$$

where $(\pi_1^\star, \pi_2^\star, \beta^\star)$ is an equilibrium solution to eqns (8.3)-(8.5), $\pi^\star = \pi_1^\star + \pi_2^\star$, $\hat{\theta} = 1-(1-\theta)(1-\phi)$ and $L = 1+(\delta_t/(\mu+\delta))$. Using eqn (8.4) we have that $\pi_1^\star = \pi^\star\eta_1$ and $\pi_2^\star = \pi^\star\eta_2$ where $\eta_1 = 1/L$ and $\eta_2 = (\delta_t/(\mu+\delta))\eta_1$. Hence eqn (8.8) becomes

$$\beta^\star = \frac{(\lambda_1\eta_1+\lambda_2\eta_2)\pi^\star}{(\lambda_1\eta_1+\lambda_2\eta_2)\pi^\star+\lambda_1(1-\pi^\star)\hat{\theta}+\hat{\tau}_1}, \tag{8.9}$$

and from eqn (8.7) we have that

$$\beta^* = \frac{\pi^*(\mu+\delta_t+\delta)}{(1-\pi^*)\lambda_1\alpha(1-\phi)L}. \tag{8.10}$$

Equating eqns (8.9) and (8.10) and simplifying we find that if $\pi^* > 0$ then

$$\pi^* = \frac{\hat{\alpha}\hat{\lambda}-\hat{\mu}(\lambda_1\hat{\theta}+\hat{\tau}_1)}{\hat{\lambda}\hat{\mu}-\hat{\mu}\lambda_1\hat{\theta}+\hat{\alpha}\hat{\lambda}} \tag{8.11}$$

$$= \frac{\hat{\alpha}\hat{\lambda}}{\hat{\lambda}\hat{\mu}-\hat{\mu}\lambda_1\hat{\theta}+\hat{\alpha}\hat{\lambda}}\left(1-\frac{\hat{\mu}(\lambda_1\hat{\theta}+\hat{\tau}_1)}{\hat{\alpha}\hat{\lambda}}\right), \tag{8.12}$$

where $\hat{\alpha} = \lambda_1\alpha(1-\phi)L$, $\hat{\lambda} = \lambda_1\eta_1+\lambda_2\eta_2$ and $\hat{\mu} = \mu+\delta_t+\delta$. Now substituting the expression for π^* in eqn (8.12) into eqn (8.10) and simplifying we have that

$$\beta^* = \frac{\hat{\alpha}\hat{\lambda}-\hat{\mu}(\lambda_1\hat{\theta}+\hat{\tau}_1)}{\hat{\alpha}(\hat{\lambda}+\hat{\tau}_1)}, \tag{8.13}$$

$$= \frac{\hat{\lambda}}{\hat{\lambda}+\hat{\tau}_1}\left(1-\frac{\hat{\mu}(\lambda_1\hat{\theta}+\hat{\tau}_1)}{\hat{\alpha}\hat{\lambda}}\right). \tag{8.14}$$

Using eqn (8.6) we have directly that

$$(\pi^*,\beta^*) = \left(\frac{\hat{\alpha}\hat{\lambda}}{\hat{\mu}\hat{\lambda}-\hat{\mu}\lambda_1\hat{\theta}+\hat{\alpha}\hat{\lambda}}\left(1-\frac{1}{R_0}\right), \frac{\hat{\lambda}}{\hat{\lambda}+\hat{\tau}_1}\left(1-\frac{1}{R_0}\right)\right). \tag{8.15}$$

Hence if $R_0 \leq 1$ then the only equilibrium solution to eqns (8.3)-(8.5) is the disease-free solution. If $R_0 > 1$ then there still exists the disease-free solution but now we also have a unique endemic equilibrium solution. This completes the equilibrium results of Theorem 8.1.•

We now show that the disease-free equilibrium is globally stable when $R_0 \leq 1$ and therefore disease is always eventually eradicated when $R_0 \leq 1$. Let $\mathbf{x} = (\pi_1, \pi_2, \beta)$ and $u(\mathbf{x}) = \beta + c_1\pi_1 + c_2\pi_2$ where

$$c_1 = \frac{\gamma\left(\frac{\tau}{\lambda_1\gamma}+\hat{\theta}\right)}{\alpha(1-\phi)} \tag{8.16}$$

and

$$c_2 = \frac{1}{\delta_t}\Big((\mu+\delta_t+\delta)c_1-\lambda_1\gamma\Big). \tag{8.17}$$

Note that $c_1 > 0$ always and when $R_0 \leq 1$, $c_2 > 0$. Hence $u(\mathbf{x})$ is positive definite on $\mathbf{x} \geq \mathbf{0}$. Using eqns (8.3)-(8.5) it is straightforward to compute du/dt, we have that

$$\frac{du}{dt} = -\pi_2\left[\frac{(\mu+\delta)(\mu+\delta_t+\delta)\gamma\left(\hat{\theta}+\frac{\tau}{\lambda_1\gamma}\right)}{\delta_t\alpha(1-\phi)}-\gamma\lambda_2-\frac{(\mu+\delta)\lambda_1\gamma}{\delta_t}\right]-\beta\pi_1(\lambda_1\gamma+\tau)$$

$$-\beta\pi_2(\lambda_2\gamma+\tau). \tag{8.18}$$

When $R_0 \leq 1$ all coefficients in du/dt are strictly negative and therefore du/dt is at least negative semi-definite on $\mathbf{x} \geq \mathbf{0}$. When $du/dt < 0$, $u(\mathbf{x})$ is a strong Lyapunov function for the disease-free solution and hence by Theorem 10.2 in Jordan and Smith (1987) the disease-free solution is globally asymptotically stable in $\mathbf{x} \geq \mathbf{0}$. We now show that when $du/dt = 0$ and $R_0 \leq 1$ the only invariant set is the disease-free solution. When $R_0 = 1$ then the π_2 coefficient in du/dt equals zero. Hence $du/dt = 0$ is satisfied when either $\beta > 0$ and $\pi = 0$ or $\beta = 0$ and $\pi > 0$, however $d\pi/dt > 0$ when $\beta > 0$ and $\pi = 0$ and $d\beta/dt > 0$ when $\beta = 0$ and $\pi > 0$. Therefore the only invariant set when $R_0 = 1$ is the disease-free solution. When $R_0 < 1$ then we must have that $\pi_2 = 0$, and either $\beta = 0$ and $\pi_1 > 0$, or $\beta > 0$ and $\pi_1 = 0$. By a similar argument to the case where $R_0 = 1$ the only invariant set in $du/dt = 0$ is the disease-free solution. By LaSalle's Invariance Principle, (La Salle, 1976), $\mathbf{x}(t) \to M \cap u^{-1}(c)$ for some $c \geq 0$ where M is the largest invariant set in $du/dt = 0$. Hence the disease-free solution is globally asymptotically stable in $\mathbf{x} \geq \mathbf{0}$.

We now show that when $R_0 > 1$ the disease-free equilibrium is no longer stable and extend this result to show that if initially present then disease will persist for all time. Consider the linearised system of eqns (8.3)-(8.5), evaluated at the disease-free equilibrium. This system can be represented in matrix form as

$$\frac{d\mathbf{x}}{dt} = \mathbf{J}\mathbf{x},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \beta)$ and

$$\mathbf{J} = \begin{bmatrix} -(\mu + \delta_t + \delta) & 0 & \lambda_1\alpha(1-\phi) \\ \delta_t & -(\mu+\delta) & 0 \\ \lambda_1\gamma & \lambda_2\gamma & -\lambda_1\gamma\hat{\theta} - \tau \end{bmatrix}.$$

We wish to show that at least one eigenvalue of $\mathbf{J}$ has a strictly positive real part. If the characteristic equation of $\mathbf{J}$ is

$$\omega^3 + a_1\omega^2 + a_2\omega + a_3 = 0, \tag{8.19}$$

then using the Routh-Hurwitz conditions we wish to show that at least one of $a_1 > 0$, $a_2 > 0$, $a_3 > 0$ and $a_1a_2 > a_3$ is not true when $R_0 > 1$. It is easy to show that

$$\begin{aligned} a_3 &= (\mu+\delta)(\mu+\delta_t+\delta)(\lambda_1\gamma\hat{\theta}+\tau) - \delta_t\lambda_1\alpha(1-\phi)\gamma\lambda_2 - \gamma\lambda_1^2\alpha(1-\phi)(\mu+\delta), \\ &= (\mu+\delta)(\mu+\delta_t+\delta)(\lambda_1\gamma\hat{\theta}+\tau)\left[1 - \frac{\lambda_1\gamma\alpha(1-\phi)[\lambda_2\delta_t + \lambda_1(\mu+\delta)]}{(\mu+\delta)(\mu+\delta_t+\delta)(\lambda_1\gamma\hat{\theta}+\tau)}\right], \\ &= (\mu+\delta)(\mu+\delta_t+\delta)(\lambda_1\gamma\hat{\theta}+\tau)(1-R_0), \end{aligned}$$

using eqn (8.6). Hence $a_3 < 0$ when $R_0 > 1$ and therefore the disease-free equilibrium is unstable when $R_0 > 1$. We can now use this result to show the following lemma:

Lemma 8.1 *There is a small (fixed) number $\epsilon > 0$ such that if $R_0 > 1$ and disease is initially present then there exists some $\nu > 0$ such that for $t \geq \nu$, $\pi_1(t) \geq \epsilon$, $\pi_2(t) \geq \epsilon$ and $\beta(t) \geq \epsilon$, where ϵ depends on the model parameters but not the initial conditions.*

Proof.

Firstly we show that if any of $\pi_1(0)$, $\pi_2(0)$ or $\beta(0)$ are strictly positive then $\pi_1(\Delta t) > 0$, $\pi_2(\Delta t) > 0$ and $\beta(\Delta t) > 0$ for Δt small and strictly positive. Following a similar method to Lemma 2.3 we need to consider four initial conditions.

1. Suppose that $\beta(0) = 0$. Hence $\pi(0) > 0$. Using eqn (8.5) we find that

$$\begin{aligned} \beta(\Delta t) &= (\pi_1(0)\lambda_1 + \pi_2(0)\lambda_2)\gamma\Delta t + o(\Delta t) > 0, \\ \text{and} \qquad \pi(\Delta t) &> 0, \qquad \text{by continuity for small } \Delta t. \end{aligned}$$

Let $\psi = 1 - \pi$, hence

$$\frac{d\psi}{dt} = -\psi\lambda_1\beta\alpha(1-\phi) + (\mu+\delta)(1-\psi).$$

If $\pi(0) < 1$ we must have $\psi(0) > 0$, hence $\psi(\Delta t) > 0$ for small enough $\Delta t > 0$. If $\pi(0) = 1$ then $\psi(0) = 0$ and

$$\psi(\Delta t) \geq (\mu+\delta)\Delta t + o(\Delta t) > 0, \text{ for } \Delta t \text{ small and strictly positive.}$$

By choosing $\Delta t > 0$ small enough and starting at $t = \Delta t$ instead of $t = 0$ we can assume that $\pi(0) > 0$, $\psi(0) > 0$ and $\beta(0) > 0$. If $\pi_1(0) = 0$ then $\pi_1(\Delta t) = \psi(0)\lambda_1\beta(0)\alpha(1-\phi)\Delta t + o(\Delta t) > 0$, if $\Delta t > 0$ is small enough. Hence we can also assume that $\pi_1(0) > 0$. If $\pi_2(0) = 0$ then $\pi_2(\Delta t) = \delta_t\pi_1(0)\Delta t + o(\Delta t) > 0$, if $\Delta t > 0$ is small enough. Therefore we can also assume that $\pi_2(0) > 0$

2. Suppose that $\pi(0) = 0$. Hence $\beta(0) > 0$. Following the same method as in the previous case we find that

$$\begin{aligned} \pi(\Delta t) &= \lambda_1\beta(0)\alpha(1-\phi)\Delta t + o(\Delta t) > 0, \qquad \text{for small } \Delta t, \\ \beta(\Delta t) &> 0, \qquad \text{by continuity for small } \Delta t, \\ \text{and} \quad \psi(\Delta t) &> 0, \qquad \text{also by continuity for small } \Delta t. \end{aligned}$$

Therefore by choosing Δt small enough and starting at $t = \Delta t$ we can assume that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$ and arguing as in the previous case we can also assume without loss of generality that $\pi_1(0) > 0$ and $\pi_2(0) > 0$.

3. Suppose that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) > 0$. This case is trivial and follows directly as in Case 1.

4. Suppose that $\pi(0) > 0$, $\beta(0) > 0$ and $\psi(0) = 0$. This implies that $\pi(0) = 1$, and hence

$$\psi(\Delta t) \geq (\mu + \delta)\Delta t + o(\Delta t) > 0.$$

Thus it follows directly that by starting at time $t = \Delta t$ where Δt is sufficiently small we can assume that $\psi(0) > 0$ and the result follows by Case 3.

Hence we can assume that if any of π_1, π_2 or β are initially strictly positive then after a short duration all components are strictly positive. Therefore we have that there exists fixed $1 > \epsilon > 0$ such that if Δt is small enough and strictly positive then $\pi_1(\Delta t) \geq \epsilon\pi_1^\star$, $\pi_2(\Delta t) \geq \epsilon\pi_2^\star$ and $\beta(\Delta t) \geq \epsilon\beta^\star$. We now show that once this occurs each component can be bounded below by a similar form of bound for all time. We must have that either $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$ or else $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$. Suppose first that $\pi_{1,\infty} \geq \frac{1}{2}\epsilon\pi_1^\star$. From the definition of $\pi_{1,\infty}$ we have that there exists T_1 such that $\pi_1 \geq \frac{1}{4}\epsilon\pi_1^\star$ for all $t \geq T_1$. From Lemma 2.2 we have directly that

$$\pi_{2,\infty} \geq \frac{\delta_t \pi_{1,\infty}}{\mu + \delta} \geq \frac{\frac{1}{2}\epsilon\delta_t\pi_1^\star}{\mu + \delta} = \frac{1}{2}\epsilon\pi_2^\star.$$

Hence there exists T_2 such that $\pi_2 \geq \frac{1}{4}\epsilon\pi_2^\star$ for all $t \geq T_2$. Using eqn (8.5) we have that

$$\frac{d\beta}{dt} \geq \lambda_1\gamma\pi_1 - \beta[\lambda_1\gamma(1 + \hat{\theta}) + \tau],$$

and following the method of Lemma 2.2 we have directly that

$$\beta_\infty \geq \frac{\lambda_1\gamma\pi_{1,\infty}}{\lambda_1\gamma(1 + \hat{\theta}) + \tau}. \tag{8.20}$$

So we have that there exists T_3 such that $\beta \geq \frac{1}{4}\epsilon_1\beta^\star$ for all $t \geq T_3$ where

$$\epsilon_1 = \frac{\epsilon\lambda_1\gamma}{\lambda_1\gamma(1 + \hat{\theta}) + \tau}\frac{\pi_1^\star}{\beta^\star}.$$

Therefore for $t \geq T = \max(T_1, T_2, T_3)$ we have that the disease persists in addicts and needles, and $\pi_1(t) \geq \epsilon_2\pi_1^\star$, $\pi_2(t) \geq \epsilon_2\pi_2^\star$ and $\beta(t) \geq \epsilon_2\beta^\star$ where $\epsilon_2 = \frac{1}{4}\min(\epsilon, \epsilon_1)$.

Now suppose that $\pi_{1,\infty} < \frac{1}{2}\epsilon\pi_1^\star$ in which case there exists $\zeta \geq \Delta t$ where $\pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star$. Let $t_0 = \inf\{\zeta \geq \Delta t, \pi_1(\zeta) < \frac{1}{2}\epsilon\pi_1^\star\}$, and $t_1 = \inf\{\zeta \geq t_0, \pi_1(\zeta) > \frac{1}{2}\epsilon\pi_1^\star\}$, where ϵ

is fixed and positive. By the definition of t_0 we have that $\pi_1(t_0+\nu) < \frac{1}{2}\epsilon\pi_1^\star$ if ν is small and positive, hence $t_1 > t_0$. Suppose that $t_1 < \infty$. By continuity $\pi_1(t_0) = \pi_1(t_1) = \frac{1}{2}\epsilon\pi_1^\star$, and therefore π_1 is less than $\frac{1}{2}\epsilon\pi_1^\star$ in (t_0, t_1) and greater than $\frac{1}{2}\epsilon\pi_1^\star$ just after t_1. t_0 is the first time after Δt that π_1 goes below $\frac{1}{2}\epsilon\pi_1^\star$, t_1 the next time after t_0 that π_1 rises above it. The basic idea is the same as in the proof of the corresponding result for the Simple Model. We shall show that π_1 being small forces π_2 and β to become small. Using the instability of the disease-free equilibrium we deduce that if ϵ is small enough then π_1, π_2 and β all being small forces π_1 to rise above $\frac{1}{2}\epsilon\pi_1^\star$, and moreover the total time for which π_1 remains continuously below $\frac{1}{2}\epsilon\pi_1^\star$ can be bounded above by a bound that depends only on ϵ and the model parameters. From Lemma 2.4 there exists a time $\bar{T}_1 > 0$ such that if $t_0 + \bar{T}_1 < t_1$ then for all $t \in [t_0 + \bar{T}_1, t_1]$, $0 < \pi_2 < (\frac{1}{2} + \Delta)\pi_2^\star\epsilon$, where Δ is small and $\bar{T}_1$ depends only on the model parameters, Δ and ϵ. We have that

$$\begin{aligned} \frac{d\beta}{dt} &\leq \lambda_1\gamma\pi - \beta\tau, \\ &\leq \lambda_1\gamma\left(\frac{1}{2} + \Delta\right)\epsilon\pi^\star - \beta\tau, \qquad \text{for } t_1 \geq t \geq t_0 + \bar{T}_1, \end{aligned}$$

and using the method of Lemma 2.4 we have directly that

$$\beta \leq \frac{\lambda_1\gamma(\frac{1}{2} + 2\Delta)\epsilon\pi^\star}{\tau} = \left(\frac{1}{2} + 2\Delta\right)\epsilon_1\beta^\star \qquad \text{for } t_1 \geq t \geq t_0 + \bar{T}_1 + \bar{T}_2,$$

where $\epsilon_1 = (\epsilon\lambda_1\gamma\pi^\star)/(\tau\beta^\star)$ and where $\bar{T}_2$ also depends only on the model parameters, Δ and ϵ. We must have that either $t_1 \geq t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2]$ or $t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2]$. We have that for $t_1 > t \geq t_0 + \bar{T}_1$ we can choose ϵ small enough such that $\frac{1}{2}\epsilon\pi_1^\star + (\frac{1}{2} + \Delta)\epsilon\pi_2^\star < \epsilon_2$ for ϵ_2 arbitrarily small and positive. Now let $t_2 = \inf\{\zeta : \text{ for } t_1 > t \geq t_0 + \zeta,\ \pi(t) < \epsilon_2\}$, and hence if $t_2 > 0$, then $\pi(t_0 + t_2) = \epsilon_2$ and $t_0 + t_2$ is the last time after t_0 but before t_1 that $\pi(t) \geq \epsilon_2$ and note that $t_2 \leq \bar{T}_1$. For $t_1 > t \geq t_0 + t_2$ using eqns (8.3)-(8.5) we find that

$$\frac{d\pi_1}{dt} \geq (1 - \epsilon_2)\lambda_1\alpha(1 - \phi)\beta - (\mu + \delta_t + \delta)\pi_1, \tag{8.21}$$

$$\frac{d\pi_2}{dt} = \delta_t\pi_1 - (\mu + \delta)\pi_2, \tag{8.22}$$

and

$$\frac{d\beta}{dt} \geq \gamma(\pi_1\lambda_1 + \pi_2\lambda_2) - \beta(\lambda_1\gamma\hat{\theta} + \tau) - (\lambda_1 + \lambda_2)\gamma\epsilon_2\beta. \tag{8.23}$$

We can write this in vector form as

$$\frac{d\mathbf{x}}{dt} \geq \mathbf{J}(\epsilon_2)\mathbf{x},$$

where $\mathbf{x}^T = (\pi_1, \pi_2, \beta)$ and

$$\mathbf{J}(\epsilon_2) = \begin{bmatrix} -(\mu+\delta_t+\delta) & 0 & \lambda\alpha(1-\phi)(1-\epsilon_2) \\ \delta_t & -(\mu+\delta) & 0 \\ \lambda_1\gamma & \lambda_2\gamma & -(\lambda_1\gamma(\epsilon_2+\hat{\theta})+\lambda_2\gamma\epsilon_2+\tau) \end{bmatrix}.$$

When $\epsilon_2 = 0$, $\mathbf{J}(0) = \mathbf{J}$, the linearised stability matrix about the disease-free equilibrium. We have already shown that when $R_0 > 1$ this equilibrium is unstable. Hence following the method of Lemma 2.5 we have that $\mathbf{J}(\epsilon_2)$ has a strictly positive left eigenvector corresponding to a strictly positive eigenvalue for ϵ_2 sufficiently small. It follows as in the proof of Lemma 2.5 that $t_1 < t_0 + \max[\bar{T}_1, \bar{T}_1 + \bar{T}_2, t_2 + \bar{T}_3]$ where $\bar{T}_3$ is a fixed time dependent only on the model parameters, ϵ_2 and Δ. Therefore arguing as in the corresponding result for the Simple Model in Chapter 2 we have that if $R_0 > 1$ and disease is initially present then there exists a fixed $\epsilon > 0$ and $\nu > 0$ such that for all $t \geq \nu$, $\pi_1(t) > \epsilon$, $\pi_2(t) > \epsilon$, and $\beta(t) > \epsilon$. Moreover ϵ and ν depend only on the model parameters. This completes the proof of Lemma 8.1.•

We have shown that if $R_0 > 1$ and disease is initially present then disease will persist indefinitely among the population. We now show that the unique endemic equilibrium is locally stable. We find that the Jacobian matrix of the system (8.3)-(8.5) evaluated at the endemic equilibrium is

$$\begin{bmatrix} -\lambda_1\alpha(1-\phi)\beta^*-(\mu+\delta_t+\delta) & -\lambda_1\alpha(1-\phi)\beta^* & (1-\pi^*)\lambda_1\alpha(1-\phi) \\ \delta_t & -(\mu+\delta) & 0 \\ (1-\beta^*+\beta^*\hat{\theta})\lambda_1\gamma & ((1-\beta^*)\lambda_2+\beta^*\lambda_1\hat{\theta})\gamma & -\gamma\Big[\lambda_1\hat{\theta}+\pi_1^*\lambda_1(1-\hat{\theta})+\pi_2^*(\lambda_2-\lambda_1\hat{\theta})\Big]-\tau \end{bmatrix}.$$

Suppose that the characteristic equation of this matrix is $\omega^3 + \bar{a}_1\omega^2 + \bar{a}_2\omega + \bar{a}_3 = 0$. We now show that the Routh-Hurwitz conditions $\bar{a}_1 > 0$, $\bar{a}_2 > 0$, $\bar{a}_3 > 0$ and $\bar{a}_1\bar{a}_2 > \bar{a}_3$ are satisfied when $R_0 > 1$. It is easy to show that

$$\bar{a}_1 = \gamma\Big[\lambda_1\hat{\theta}(1-\pi^*) + \pi_1^*\lambda_1 + \pi_2^*\lambda_2\Big] + \tau + \lambda_1\alpha(1-\phi)\beta^* + (\mu+\delta_t+\delta) + \mu + \delta,$$

and hence $\bar{a}_1 > 0$. We find that

$$\begin{aligned} \bar{a}_2 = {} & (\lambda_1\alpha(1-\phi)\beta^* + \mu + \delta_t + \delta)\gamma\Big[\lambda_1\hat{\theta}(1-\pi^*) + \pi_1^*\lambda_1 + \pi_2^*\lambda_2 + \hat{\tau}_1\Big] \\ & + \gamma\Big[\lambda_1\hat{\theta}(1-\pi^*) + \pi_1^*\lambda_1 + \pi_2^*\lambda_2 + \hat{\tau}_1\Big](\mu+\delta) \\ & + (\lambda_1\alpha(1-\phi)\beta^* + \mu + \delta_t + \delta)(\mu+\delta) + \lambda_1\alpha(1-\phi)\beta^*\delta_t \\ & - \lambda_1\gamma(1-\beta^*(1-\hat{\theta}))(1-\pi^*)\lambda_1\alpha(1-\phi). \end{aligned} \tag{8.24}$$

We wish to express the positive terms in such a way as to eliminate the negative term. First note that we can cancel out the $-\lambda_1\gamma\beta^\star\hat{\theta}(1-\pi^\star)\lambda_1\alpha(1-\phi)$ term leaving us with $-\lambda_1\gamma(1-\beta^\star)(1-\pi^\star)\lambda_1\alpha(1-\phi)$. We have that

$$-\lambda_1\gamma(1-\beta^\star)(1-\pi^\star)\lambda_1\alpha(1-\phi) = -\lambda_1\gamma\frac{(1-\beta^\star)}{\beta^\star}\pi^\star\eta_1(\mu+\delta_t+\delta),$$

using eqn (8.7),

$$= -\lambda_1\gamma\pi^\star\eta_1(\mu+\delta_t+\delta)\frac{[\lambda_1\gamma(1-\pi^\star)\hat{\theta}+\tau]}{\gamma\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)},$$

using eqn (8.8),

$$= \frac{-\lambda_1\eta_1}{\lambda_1\eta_1+\lambda_2\eta_2}(\mu+\delta_t+\delta)[\lambda_1\gamma(1-\pi^\star)\hat{\theta}+\tau].$$

In $\bar{a}_2$ we have a positive term $(\mu+\delta_t+\delta)[\gamma\lambda_1\hat{\theta}(1-\pi^\star)+\tau]$ and since $\lambda_1\eta_1/(\lambda_1\eta_1+\lambda_2\eta_2) < 1$ then it follows directly that $\bar{a}_2 > 0$. We now move on to $\bar{a}_3$, we have that

$$\begin{aligned}
\bar{a}_3 = {} & (\lambda_1\alpha(1-\phi)\beta^\star+\mu+\delta_t+\delta)(\mu+\delta)\left[\lambda_1\hat{\theta}(1-\pi^\star)+\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)+\hat{\tau}_1\right]\gamma \\
& +\delta_t\lambda_1\alpha(1-\phi)\beta^\star\gamma\left[\lambda_1\hat{\theta}(1-\pi^\star)+\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)+\hat{\tau}_1\right] \\
& -\delta_t(1-\pi^\star)\lambda_1\alpha(1-\phi)\gamma[\lambda_2(1-\beta^\star)+\beta^\star\lambda_1\hat{\theta}] \\
& -\lambda_1\gamma(1-\beta^\star(1-\hat{\theta}))(1-\pi^\star)\lambda_1\alpha(1-\phi)(\mu+\delta).
\end{aligned}$$

This expression for $\bar{a}_3$ simplifies down to

$$\begin{aligned}
\bar{a}_3 = {} & \lambda_1\alpha(1-\phi)\beta^\star(\mu+\delta)\gamma[\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)+\hat{\tau}_1] \\
& +(\mu+\delta_t+\delta)(\mu+\delta)\left[\lambda_1\hat{\theta}(1-\pi^\star)+\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)+\hat{\tau}_1\right]\gamma \\
& +\delta_t\lambda_1\alpha(1-\phi)\beta^\star\gamma[\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)+\hat{\tau}_1] \\
& -(1-\beta^\star)(1-\pi^\star)\lambda_1\alpha(1-\phi)\gamma[\delta_t\lambda_2+(\mu+\delta)\lambda_1].
\end{aligned}$$

In a similar fashion to the $\bar{a}_2 > 0$ calculation we find that

$$-(1-\beta^\star)(1-\pi^\star)\lambda_1\alpha(1-\phi)\gamma[\delta_t\lambda_2+(\mu+\delta)\lambda_1]$$

$$= -(\mu+\delta_t+\delta)(\mu+\delta)[\lambda_1\gamma(1-\pi^\star)\hat{\theta}+\tau].$$

Hence we can cancel out the negative term in $\bar{a}_3$ using part of

$$(\mu+\delta_t+\delta)(\mu+\delta)\left[\lambda_1\hat{\theta}(1-\pi^\star)+\hat{\tau}_1+\pi^\star(\lambda_1\eta_1+\lambda_2\eta_2)\right]\gamma,$$

and therefore $\bar{a}_3 > 0$. We now show the condition of $\bar{a}_1\bar{a}_2 > \bar{a}_3$. This follows a similar method to the previous manipulations but contains more complicated terms. $\bar{a}_1\bar{a}_2 - \bar{a}_3 =$

$$(\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)\gamma^2\Big[\lambda_1\hat{\theta}(1-\pi^\star) + \pi^\star(\lambda_1\eta_1 + \lambda_2\eta_2) + \hat{\tau}_1\Big]^2 \tag{8.25}$$

$$+ \quad (\mu+\delta)\gamma^2\Big[\lambda_1\hat{\theta}(1-\pi^\star) + \pi^\star(\lambda_1\eta_1 + \lambda_2\eta_2) + \hat{\tau}_1\Big]^2$$

$$+ \quad (\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)^2\gamma\Big[\lambda_1\hat{\theta}(1-\pi^\star) + \pi^\star(\lambda_1\eta_1 + \lambda_2\eta_2) + \hat{\tau}_1\Big] \tag{8.26}$$

$$+ \quad 2(\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)(\mu+\delta)\Big[\lambda_1\hat{\theta}(1-\pi^\star) + \pi^\star(\lambda_1\eta_1 + \lambda_2\eta_2) + \hat{\tau}_1\Big]\gamma$$

$$+ \quad (\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)^2(\mu+\delta)$$

$$+ \quad (\mu+\delta)^2\gamma\Big[\lambda_1\hat{\theta}(1-\pi^\star) + \pi^\star(\lambda_1\eta_1 + \lambda_2\eta_2) + \hat{\tau}_1\Big]$$

$$+ \quad (\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)(\mu+\delta)^2$$

$$- \quad \lambda_1\gamma^2(1-\beta^\star(1-\hat{\theta}))(1-\pi^\star)\lambda_1\alpha(1-\phi)\Big[\lambda_1\hat{\theta}(1-\pi^\star) + \pi^\star(\lambda_1\eta_1 + \lambda_2\eta_2) + \hat{\tau}_1\Big] \tag{8.27}$$

$$+ \quad (\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)\delta_t\lambda_1\alpha(1-\phi)\beta^\star$$

$$- \quad (\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)\lambda_1\gamma(1-\beta^\star(1-\hat{\theta}))(1-\pi^\star)\lambda_1\alpha(1-\phi) \tag{8.28}$$

$$+ \quad (\mu+\delta)\delta_t\lambda_1\alpha(1-\phi)\beta^\star + \delta_t(1-\pi^\star)\lambda_1\alpha(1-\phi)\gamma\Big[\lambda_2(1-\beta^\star) + \beta^\star\lambda_1\hat{\theta}\Big]. \tag{8.29}$$

We now collect the terms in expressions (8.25) and (8.27) and the terms in expressions (8.26) and (8.28), this gives us that $\bar{a}_1\bar{a}_2 - \bar{a}_3 =$

$$\Big(\lambda_1\gamma\hat{\theta}(1-\pi^\star) + \pi^\star\gamma(\lambda_1\eta_1 + \lambda_2\eta_2) + \tau\Big)\Big\{ - \lambda_1\gamma(1-\beta^\star(1-\hat{\theta}))(1-\pi^\star)\lambda_1\alpha(1-\phi)$$

$$+(\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta)\Big(\lambda_1\gamma\hat{\theta}(1-\pi^\star) + \pi^\star\gamma(\lambda_1\eta_1 + \lambda_2\eta_2) + \tau\Big)\Big\}$$

$$+\Big(\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta\Big)\Big\{ - \lambda_1\gamma(1-\beta^\star(1-\hat{\theta}))(1-\pi^\star)\lambda_1\alpha(1-\phi)$$

$$+\Big(\lambda_1\alpha(1-\phi)\beta^\star + \mu + \delta_t + \delta\Big)\Big(\lambda_1\gamma\hat{\theta}(1-\pi^\star) + \pi^\star\gamma(\lambda_1\eta_1 + \lambda_2\eta_2) + \tau\Big)\Big\}$$

$+$ other strictly positive terms.

Consider the terms inside the $\{\ldots\}$ pair, we can cancel out the $-\lambda_1\gamma\beta^*\hat{\theta}(1-\pi^*)\lambda_1\alpha(1-\phi)$ term leaving us with $-\lambda_1\gamma(1-\beta^*)(1-\pi^*)\lambda_1\alpha(1-\phi)$ plus other positive terms. We have already shown that

$$\lambda_1\gamma(1-\beta^*)(1-\pi^*)\lambda_1\alpha(1-\phi) < (\mu+\delta_t+\delta)[\lambda_1\gamma\hat{\theta}(1-\pi^*)+\tau],$$

and since $(\mu+\delta_t+\delta)[\lambda_1\gamma\hat{\theta}(1-\pi^*)+\tau]$ is an unused part of the positive terms inside $\{\ldots\}$ we have that $\bar{a}_1\bar{a}_2 > \bar{a}_3$. This completes the local stability results for the endemic equilibrium and the proof of Theorem 8.1.•

8.4.1 Interpretation of R_0

We find that there exists a critical threshold parameter which governs whether or not disease dies out or takes off in our HIV Test Model. We now show that as usual this critical threshold parameter represents the basic reproductive number. Consider a single newly infected addict entering a population containing only susceptible addicts and uninfected needles. On average a single infectious addict remains in the population unaware of his or her infectious status for $1/(\mu+\delta_t+\delta)$ time units and during this period shares needles at rate λ_1 per unit time. Hence $\lambda_1/(\mu+\delta_t+\delta)$ is the expected number of needles that are left infectious prior to the addict being tested for HIV. The probability that an addict receives an HIV test during his or her infectious lifetime is $\delta_t/(\mu+\delta_t+\delta)$. Hence an addict becomes aware of his or her HIV status with probability $\delta_t/(\mu+\delta_t+\delta)$ and now shares needles at rate λ_2 per unit time for the remainder of his or her infectious lifetime. On average a single infectious addict remains in the population after an HIV test for $1/(\mu+\delta)$ time units. Therefore $\lambda_2/(\mu+\delta)$ is the expected number of needles that are left infectious by the addict once the addict has tested positive for HIV. Hence the total expected number of needles a single infectious addict leaves infectious during his or her infectious lifetime is

$$\frac{1}{\mu+\delta_t+\delta}\left[\lambda_1+\frac{\lambda_2\delta_t}{\mu+\delta}\right].$$

We now know how many needles an addict infects during his or her infectious lifetime, we next wish to determine how many infections are caused by each needle until it is rendered virus free. Consider a single infectious needle, this needle is rendered virus free if a susceptible addict flushes or cleans the needle. In addition a needle can be rendered virus free by being exchanged. Hence the total rate at which a single infectious

needle is rendered virus free is $\lambda_1\gamma(1-(1-\phi)(1-\theta))+\tau$, therefore an infected needle spends on average

$$\frac{1}{\lambda_1\gamma(1-(1-\phi)(1-\theta))+\tau},$$

time units in an infectious state. Hence the expected number of susceptible addicts infected by this single infectious needle is

$$\frac{\lambda_1\gamma\alpha(1-\phi)}{\lambda_1\gamma(1-(1-\phi)(1-\theta))+\tau}.$$

We now have the expected number of addicts infected by a single infectious needle, putting this together with the expected number of needles an addict infects throughout his or her lifetime gives R_0. Hence,

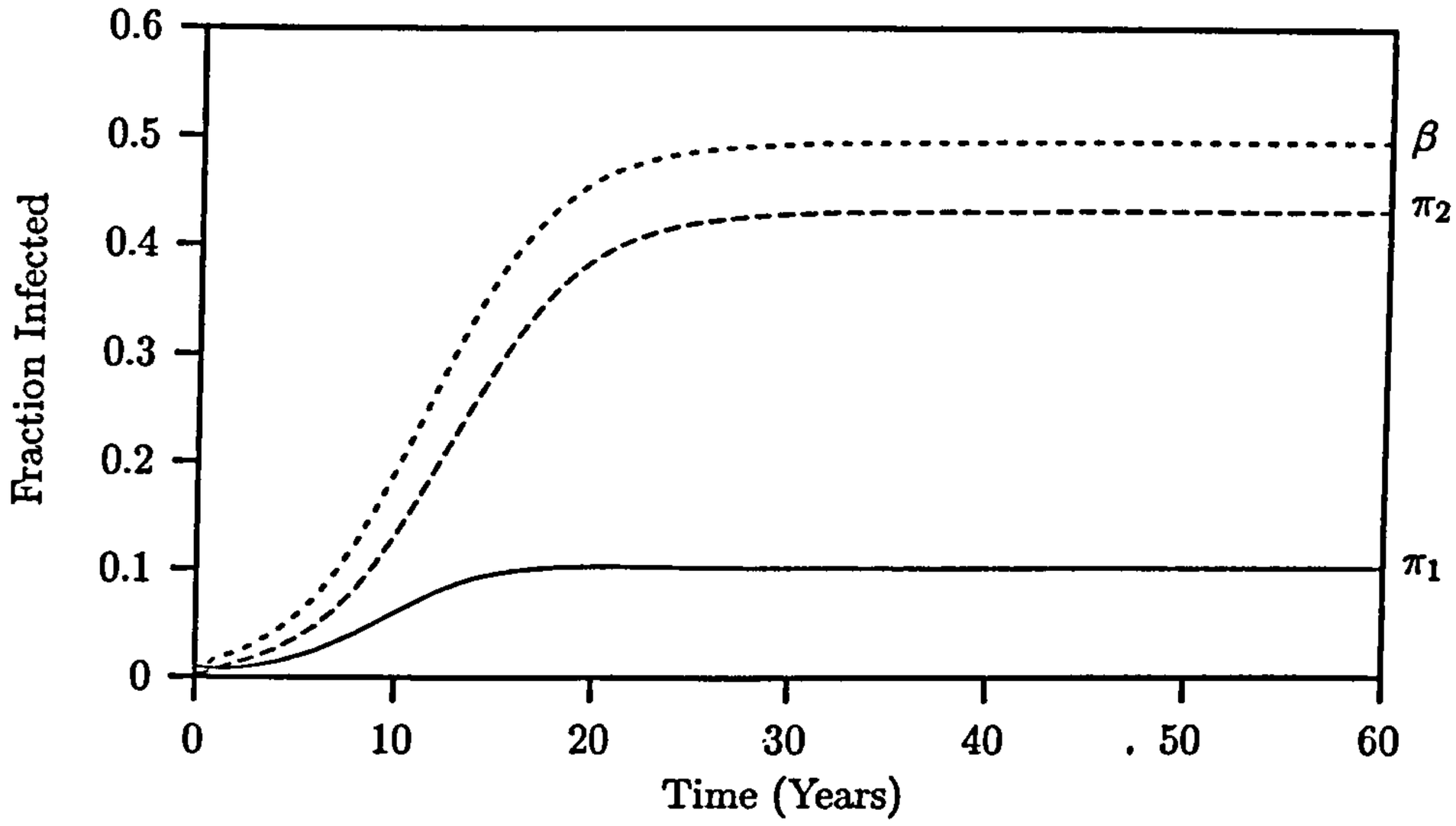
$$R_0 = \frac{\lambda_1\alpha(1-\phi)}{(\mu+\delta_t+\delta)(\hat{\tau}_1+\lambda_1\hat{\theta})}\left[\lambda_1+\frac{\lambda_2\delta_t}{\mu+\delta}\right], \tag{8.30}$$

and therefore we have that the critical threshold parameter in Theorem 8.1 is indeed the basic reproductive number. As usual R_0 also has an alternative interpretation as the expected number of secondary needles infected by a single infected needle entering an entirely susceptible population at the disease-free equilibrium.

8.5 Stability using Simulation

Before we can simulate the HIV Test Model we first need to estimate values for the parameters in this model. Where appropriate we again use the parameter estimates discussed in Appendix B and illustrated in Table 2.1, we additionally need to estimate λ_1, λ_2 and δ_t. Whatever values we use for λ_1 and λ_2 it seems intuitive that $\lambda_1 > \lambda_2$, as addicts who know they are infected should reduce the rate at which they share needles. Kaplan and O'Keefe (1993) estimate that the mean shared injection rate for the population of intravenous drug users in New Haven, Connecticut, USA, is approximately 246.22 per year. As discussed in Appendix B this estimate is higher than that suggested by other authors such as Goldberg et al. (1995) in Greenhalgh (1996) who estimate that addicts in Glasgow, Scotland, share needles only 72 times a year on average. However this distribution of needle sharing is very skew with many addicts sharing infrequently and a small minority who share equipment up to between 900 and 1800 times a year. It could be argued that it makes sense to overestimate rather than underestimate this parameter since a small minority of very high risk users could have a disproportionally large effect on the spread of the disease. We therefore

Figure 8.1: HIV Test Model Tends to Endemic Equilibrium when $R_0 > 1$

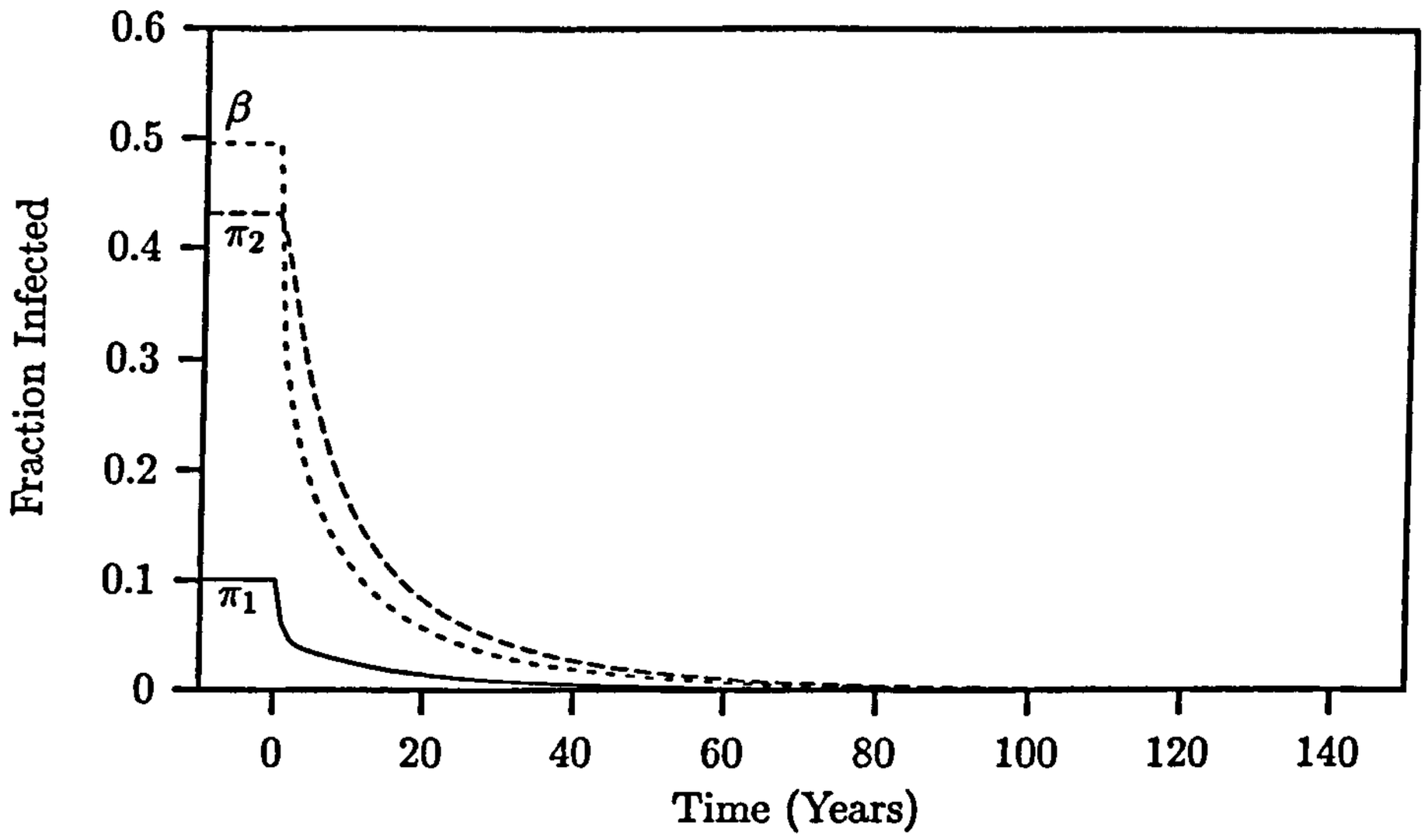


suppose that addicts who have not yet tested positive for HIV inject with a shared needle 250 times per year on average. Kretzschmar and Wiessing (1997) discuss a drop of 50% in the shared injection rate once an addict has received a positive HIV test. In the following simulations we assume that once aware of their infectious status, addicts have a mean shared injection rate of 150 per year. The Center for Disease Control in the United States estimates that if anonymous HIV testing is available then individuals who consider themselves at risk will participate in an HIV test at least once a year (CDC, 1998), therefore we assume that $\delta_t = 1$ per year.

Figure 8.1 shows simulations of the fraction of infected addicts who are not yet aware of their infectious status, the fraction of infected addicts who have tested positive for HIV and the fraction of infectious needles circulating among the total addict population. These simulations initially assume that a fraction 0.01 of all addicts are infectious where none of these addicts have yet tested positive for HIV and no other addicts or needles are initially infectious. In these simulations we assumed that $\lambda_1 = 250$ per year, $\lambda_2 = 150$ per year and $\delta_t = 1$ per year (hence addicts are tested for HIV on average once a year). The value of R_0 for these parameter estimates is 2.36 and the endemic equilibrium values are $(\pi_1^\star, \pi_2^\star, \beta^\star) = (0.101, 0.431, 0.495)$. The simulations clearly show that after approximately 30 years the model reaches a steady state corresponding to these equilibrium values.

Figure 8.2 again shows simulations of the fraction of infected addicts who are not yet

Figure 8.2: HIV Test Model Tends to Disease-Free Equilibrium when $R_0 < 1$



aware of their infectious status, the fraction of infected addicts who have tested positive for HIV and the fraction of infectious needles circulating among the total addict population. These simulations initially assume that $(\pi_1(0), \pi_2(0), \beta(0)) = (0.101, 0.431, 0.495)$, the endemic equilibrium values for the parameter estimates used in Figure 8.1. At time zero we have used the following parameter values: $\lambda_1 = 150$; $\lambda_2 = 50$; $\delta_t = 1$, these give an R_0 value of 0.84. It is clear from the simulations that eventually disease dies out in all addicts and all needles.

We simulated the HIV Test Model for a variety of different initial conditions and parameter estimates. In each case we found that if $R_0 < 1$ then disease eventually dies out in all addicts and all needles. When disease was initially present and $R_0 > 1$ then in each case the model tended to the endemic equilibrium solution.

8.6 The HIV Test Model and the Greenhalgh and Hay Model

We now examine the differences in behaviour between our HIV Test Model and the model discussed in Greenhalgh and Hay (1997). It has been demonstrated that in both these models if $R_0 \leq 1$ then the disease will die out in all addicts and all needles and if $R_0 > 1$ then the prevalence of disease appears to tend to the unique endemic equilibrium solution. We are interested in whether allowing addicts to be tested for

HIV at a constant rate per unit time (at which point addicts then change their needle sharing rate from λ_1 to λ_2) gives rise to a different long term prevalence of disease compared to the simpler situation where it is assumed that a constant proportion p of infectious addicts inject at rate λ_2 and a constant proportion $1-p$ of infectious addicts inject at rate λ_1. We now compare the endemic equilibrium solution of the HIV Test Model with that of the Greenhalgh and Hay Model. As in Chapter 7 we cannot sensibly compare these models without first making some adjustments such that we have a fair comparison.

8.6.1 HIV Model Calibration

We wish to adjust our two models such that the only difference between them is that in one, the proportion of addicts who inject at rate λ_2 is constant, and in the other the proportion of addicts who inject at rate λ_2 varies as the epidemic spreads among the addict population. In the Greenhalgh and Hay Model a fraction p of all addicts inject at rate λ_2 and a fraction $1-p$ inject at rate λ_1. In the HIV Test Model each addict spends on average

$$\frac{1}{\mu+\delta_t+\delta}$$

time units injecting at rate λ_1 and on average

$$\frac{\delta_t}{\mu+\delta_t+\delta}\frac{1}{\mu+\delta}$$

time units injecting at rate λ_2. At equilibrium the expected number of addicts in each of these two categories is inversely proportional to the time spent in each category, hence the ratio of type two to type one addicts is in the ratio

$$\frac{\delta_t}{\mu+\delta}:1.$$

Hence one way to calibrate these two models is to require the same ratio of addicts in each of these two categories for both models so

$$\frac{p}{1-p}=\frac{\delta_t}{\mu+\delta}.$$

Solving for p gives,

$$p=\frac{\delta_t}{\mu+\delta_t+\delta}.$$

It is natural to set $p=\delta_t/(\mu+\delta_t+\delta)$ since this represents the long term proportion of addicts in the HIV Test Model who share needles at rate λ_2. Hence with this choice of

p and δ

$$\frac{1-p}{\mu+\delta} = \frac{1}{\mu+\delta_t+\delta}, \tag{8.31}$$

and

$$\frac{p}{\mu+\delta} = \frac{\delta_t}{\mu+\delta_t+\delta}\frac{1}{\mu+\delta}. \tag{8.32}$$

Theorem 8.2 *Under calibration the HIV Test Model has the same endemic equilibrium solution as the model due to Greenhalgh and Hay (1997).*

Proof.

We first rewrite eqns (8.1)-(8.2) as

$$\frac{d\beta}{dt} = \pi\sigma(1-\beta) - (1-\pi)\rho\beta - \tau\beta, \tag{8.33}$$

$$\frac{d\pi}{dt} = (1-\pi)\nu\beta - (\mu+\delta)\pi, \tag{8.34}$$

where

$$\nu = \lambda_1(1-\phi)\alpha, \tag{8.35}$$

$$\rho = \lambda_1\gamma\hat{\theta}, \tag{8.36}$$

and

$$\sigma = \gamma\Big[(1-p)\lambda_1 + p\lambda_2\Big]. \tag{8.37}$$

From eqn (8.34) and eqn (8.33) we have that

$$\beta^\star = \frac{(\mu+\delta)\pi^\star}{(1-\pi^\star)\nu}, \tag{8.38}$$

and

$$\beta^\star = \frac{\sigma\pi^\star}{\rho(1-\pi^\star)+\tau+\sigma\pi^\star}. \tag{8.39}$$

Equating eqns (8.38) and (8.39) and simplifying gives us

$$\pi^\star = \frac{\nu\sigma}{(\mu+\delta)(\sigma-\rho)+\sigma\nu}\left[1-\frac{1}{R_0^{GH}}\right], \tag{8.40}$$

where

$$R_0^{GH} = \frac{\lambda_1(1-\phi)\alpha[(1-p)\lambda_1+p\lambda_2]\gamma}{(\mu+\delta)(\lambda_1\gamma\hat{\theta}+\tau)}, \tag{8.41}$$

is the basic reproductive number of the Greenhalgh and Hay Model. Substituting in the expression for $\pi^\star$ in eqn (8.40) into eqn (8.38) gives us

$$\beta^\star = \frac{\sigma}{\sigma+\tau}\left[1-\frac{1}{R_0^{GH}}\right]. \tag{8.42}$$

Using eqns (8.31) and (8.32) we have that $R_0^{GH} = R_0$ where R_0 is the basic reproductive number of the HIV Test Model. After a little simplification we find that under calibration

$$\frac{\nu\sigma}{(\mu+\delta)(\sigma-\rho)+\sigma\nu} = \frac{\hat{\alpha}\hat{\lambda}}{\hat{\mu}(\hat{\lambda}-\lambda_1\hat{\theta})+\hat{\alpha}\hat{\lambda}}, \tag{8.43}$$

and

$$\frac{\sigma}{\sigma+\tau} = \frac{\hat{\lambda}}{\hat{\lambda}+\hat{\tau}_1}. \tag{8.44}$$

Hence the result follows directly from eqn (8.15). This completes the proof of Theorem 8.2.•

We now simulate the Greenhalgh and Hay Model with p and δ_t chosen as described previously and compare these simulations with the HIV Test Model such that these models are calibrated. Figures 8.3-8.5 show comparisons of the total fraction of infected addicts, the total fraction of infected needles and the fraction of infectious addicts who have been tested for HIV. It is clear from the figures that as we expect from Theorem 8.2 the long term prevalence of disease in addicts is the same in the HIV Test Model as in the Greenhalgh and Hay Model and similarly for the long term prevalence of disease in needles. However note that the speed at which the disease spreads is less when HIV testing is viewed as a static process. An explanation for this is offered in Figure 8.5 which suggests that at the start of an epidemic the static assumption greatly overestimates the number of infectious addicts who have tested positive for HIV. However Figure 8.5 also shows that eventually the proportion of infectious addicts who have been tested for HIV reaches a steady state value which is the same as that of the constant p in the Greenhalgh and Hay Model.

To summarise, we have that under calibration the proportion of infectious addicts who have been tested for HIV and therefore inject at rate λ_2 is eventually the same in the Greenhalgh and Hay Model and the HIV Test Model. When this is the case we have that the endemic equilibrium solution in both models is the same and therefore we can conclude that modelling HIV testing by allowing addicts to flow from one class into the other, although it is more realistic, makes no quantitative long-term difference to the results of the simpler method used by Greenhalgh and Hay.

Figure 8.3: Infectious Addict Comparison

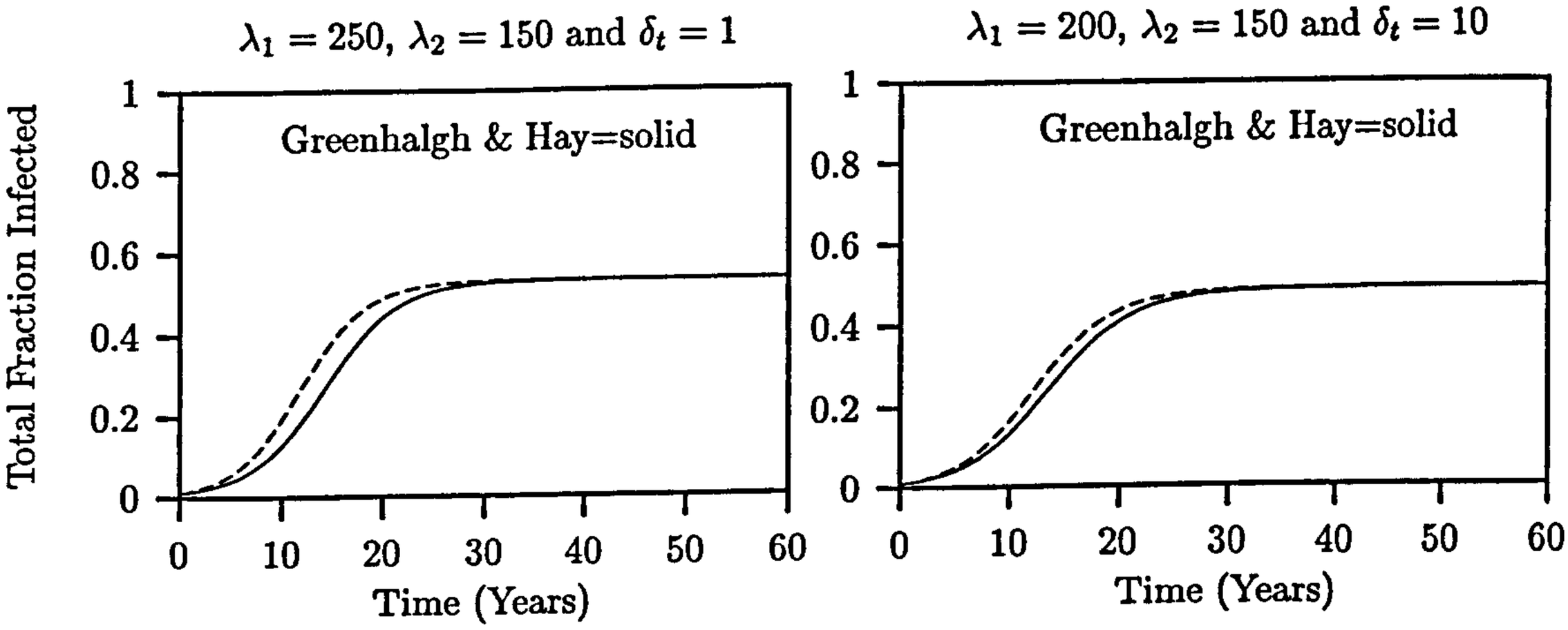


Figure 8.4: Infectious Needle Comparison

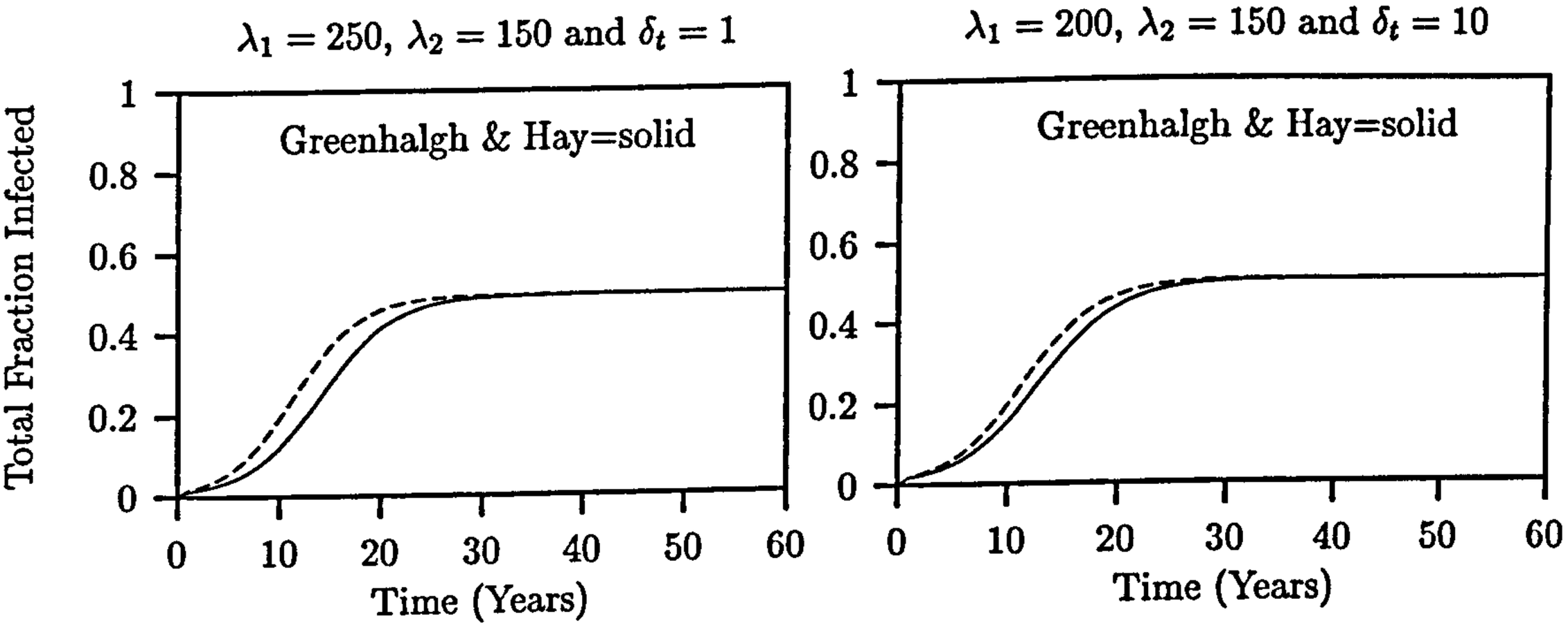
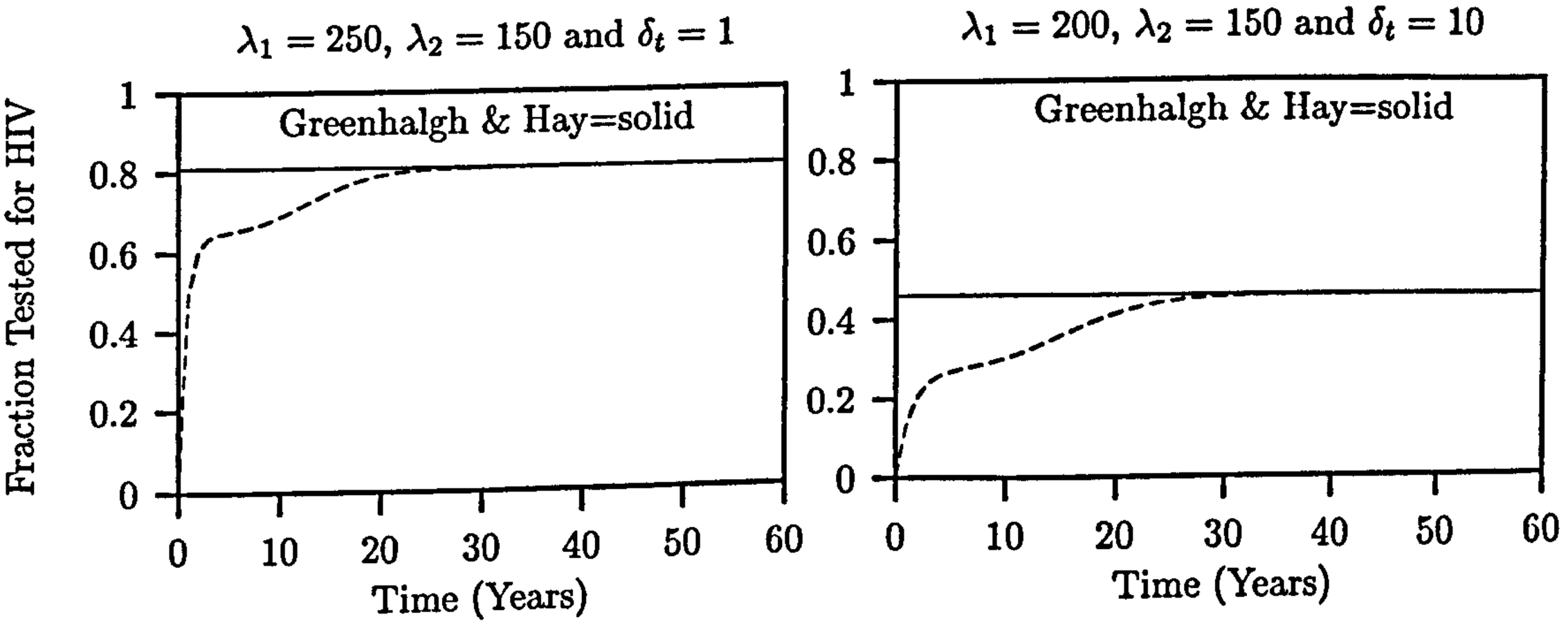


Figure 8.5: Proportion of Infectious Addicts Tested for HIV



8.7 Performance Measures of HIV Testing

We now briefly examine the effect on the long term prevalence of disease in addicts caused by introducing an HIV testing program. It was decided to examine the effect on $\pi^\star$ for $\lambda_1 = 250$ and λ_2 in the range of 250 to 50 shared injections per year and where addicts are tested for HIV from between once every ten years to once every year (on average). These estimates are consistent with those in Kretzschmar and Wiessing (1998) and CDC (1998).

Table 8.1 shows the total long term prevalence of HIV in addicts suggested by

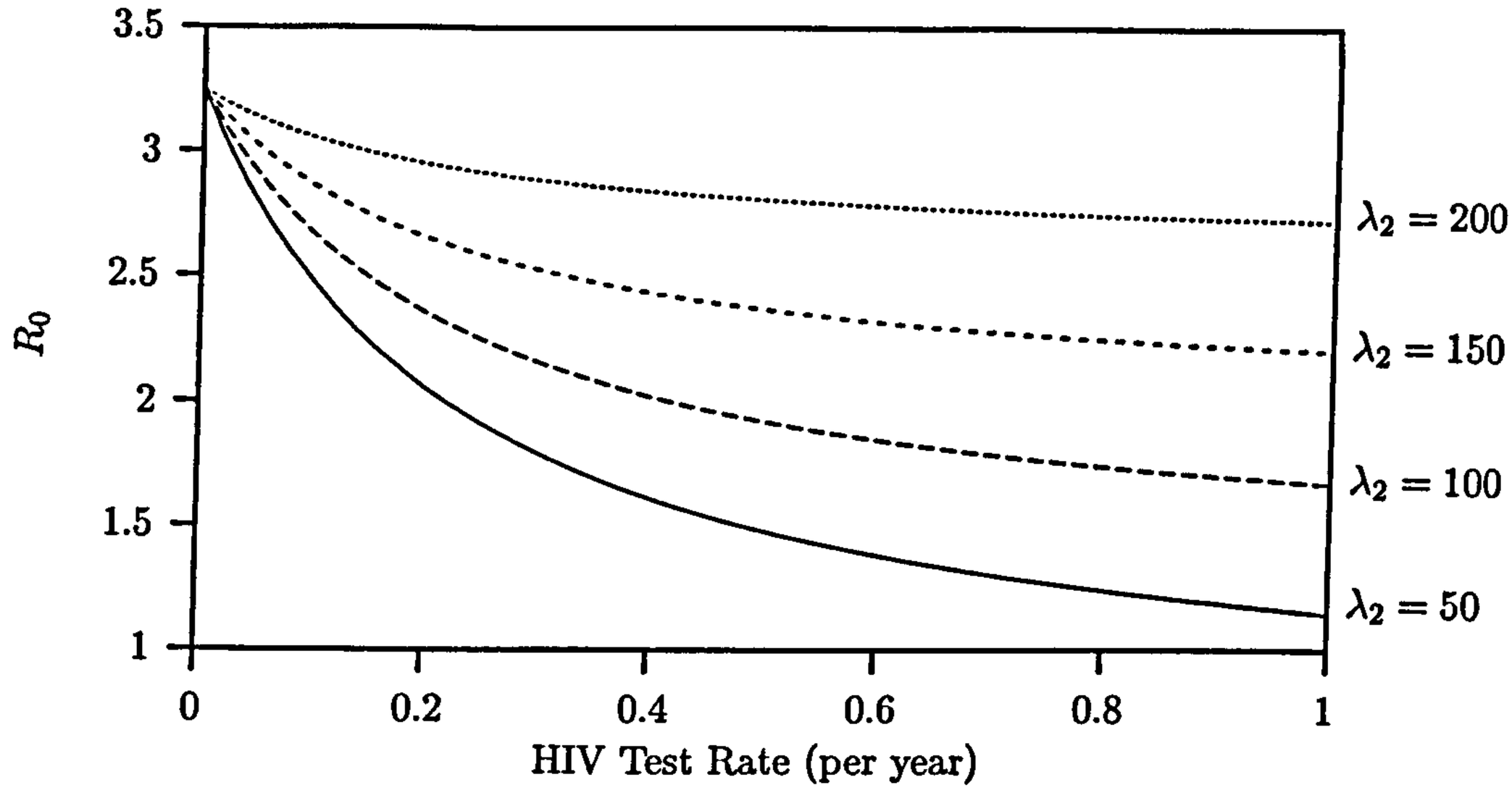
Table 8.1: Long Term Prevalence in Addicts

Injection Rate (λ_2)	HIV Testing Rate (δ_t)					
	0.1	0.2	0.3	0.4	0.5	1.0
250	59.8%	59.8%	59.8%	59.8%	59.8%	59.8%
200	59.1%	58.5%	58.3%	58.0%	57.8%	57.4%
150	58.1%	56.9%	56.0%	55.4%	54.8%	53.2%
100	57.0%	54.7%	52.8%	51.1%	49.7%	44.9%
50	55.7%	51.6%	47.5%	43.4%	39.3%	19.5%

the HIV Test Model for the range of parameters mentioned above. There are two contrasting points of interest in this table. Firstly note the very small effect caused by testing addicts if they are tested less than once a year and after being made aware of their infectious status still inject with shared needles more than 100 times a year. In contrast note the very significant decrease in prevalence when addicts are tested on average once every two years or once every year and only share needles up to 50 times a year. Therefore a fair summary of the effect of testing addicts for HIV is that this strategy will be effective only provided that addicts are tested regularly for HIV, say once a year, and moreover those addicts who have received a positive test must be encouraged to greatly reduce the rate at which they inject with shared needles.

We have previously shown that the basic reproductive number is fundamental in determining whether the disease becomes endemic or dies out. Moreover using eqn (8.15) together with simulations it is clear that R_0 also has a significant influence on the initial speed at which disease spreads and the long term prevalence of disease. Hence R_0 can also be used as a performance measure of the effectiveness of HIV testing. Figure

Figure 8.6: R_0 as a Performance Measure



8.6 illustrates R_0 as a function of the HIV testing rate for values of λ_2 from 200 shared injections per year to 50 shared injections per year. All other parameter estimates are as previously stated. The figure shows a similar behaviour to that suggested in Table 8.1 in that R_0 only becomes close to one (the critical threshold point) if both addicts are tested frequently and the post-test shared injection rate is very low.

Figures 8.7 and 8.8 demonstrate how R_0 and the long term prevalence of HIV among addicts respond to changes in needle exchange rate, (the remaining parameter estimates are as in Figure 8.1). Kaplan (1995) estimates that prior to the introduction of a formal needle exchange program, needles are exchanged 15.53 times a year on average. As discussed in Chapter 7 once a formal program has become established he estimates that this will increase to between 73 and 121.7 times a year on average. From the figures it is clear that once the needle exchange rate becomes sufficiently large disease no longer remains endemic among the population. We have that

$$R_0 < 1 \text{ if and only if } \tau > \tau_0 = \frac{\lambda_1 \alpha (1-\phi)}{\mu + \delta_t + \delta} \left[\lambda_1 + \frac{\lambda_2 \delta_t}{\mu + \delta} \right] \gamma - \lambda_1 \hat{\theta} \gamma.$$

For the parameter estimates used in the figures we have that if $\tau > 207.8$ per year then the disease will die out. It seems reasonable to expect that the reduction in the spread of HIV due to implementing both HIV testing and needle exchange together will be at least equal to (and possibly much more) than using either control measure on its own.

Evidence from our HIV Test Model suggests that HIV testing may work well as a control strategy but only if addicts are both tested regularly and significantly reduce

Figure 8.7: Effect of Needle Exchange on R_0

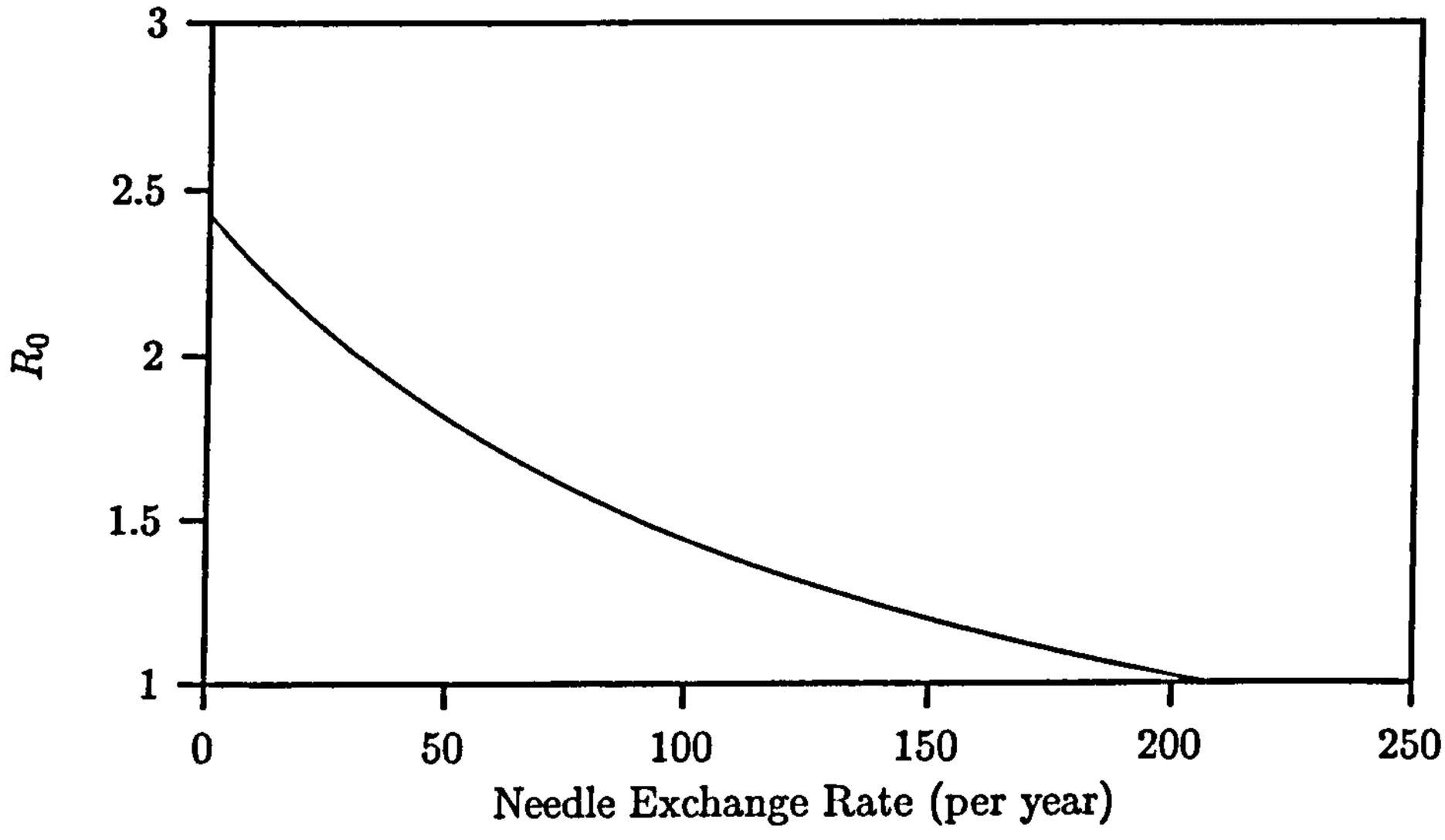
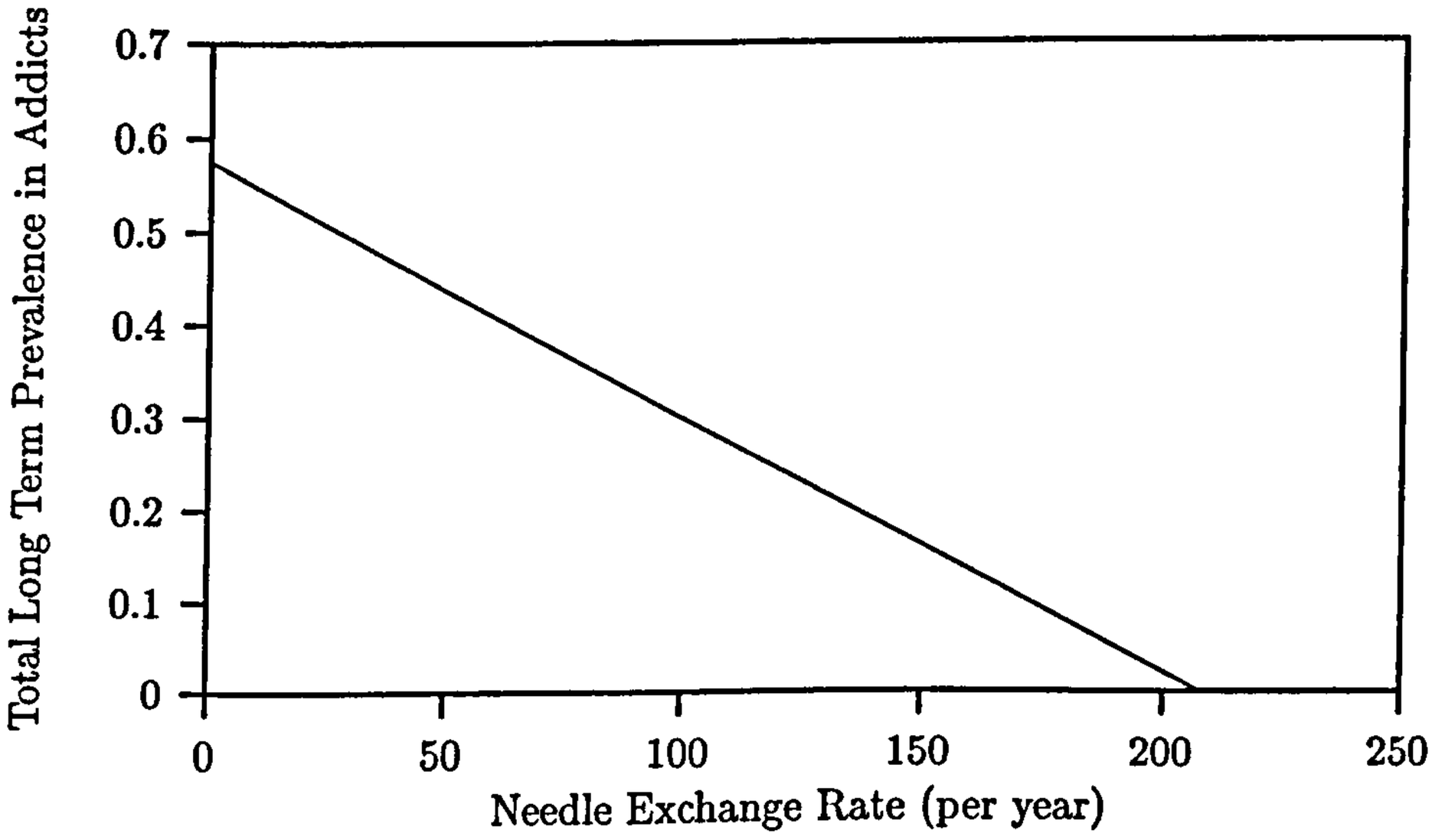


Figure 8.8: Effect of Needle Exchange on π^*



their needle sharing rate after a positive HIV test. It would seem that the most effective method of combating the spread of HIV among an intravenous drug using population would be to combine both needle exchange and HIV testing into one cohesive scheme. This would make sense since a needle exchange scheme (as described by Kaplan and O'Keefe, 1993) involves addicts regularly attending an outreach centre or something of a similar nature where addicts can exchange needles and receive counselling and advice on their risk taking behaviour. A needle exchange scheme also keeps a track of who has attended the scheme and at what date they exchanged needles. It is easy to see that this information could be useful in conjunction with an HIV testing scheme so that addicts could be tested on a regular basis. It also makes sense from an economic and practical point of view to run HIV testing from the same premises as a needle exchange scheme.

8.8 Three Stage Infectivity and HIV Testing

We now examine the impact of testing addicts for HIV when the infectivity of an addict varies throughout his or her infectious lifetime. Kretzschmar and Wiessing (1998) investigate the effect of HIV testing in a model which assumes that addicts progress through two stages of infectivity prior to the development of full blown AIDS. They assume that after infection addicts initially enter a brief period of only 60 days duration, during which time sharing a needle with this addict carries a 50% chance of HIV infection. After this stage it is assumed that for the remainder of their infectious lifetime (approximately 10 years) sharing a needle with this addict carries only a 1% chance of infection. Under these assumptions Kretzschmar and Wiessing find that HIV testing is of very little benefit. An obvious reason for this is that by the time the vast majority of infectious addicts are tested they have already entered the very low infectivity phase, and as such are of relatively little importance in causing new infections. If Kretzschmar and Wiessing had used a three stage infectious period where addict's infectivity increases prior to their departure from the population (as suggested by Peterson et al., 1990) then the effect of HIV testing may have been very different.

Our main objective of this section is to examine the effect of HIV testing using a variety of different infectivity assumptions. To do this we use the Optimistic and Pessimistic Models discussed in Chapters 3 and 4 extended to cater for HIV testing in a similar fashion to the model discussed in Section 8.3. There is a case for using instead

the General Model extended to include HIV testing as this is a more flexible model, however the drawback with this is that we do not know or have any way of estimating realistic addict-needle interaction assumptions. Instead by using the Optimistic and Pessimistic Models, we can get an idea of the best and worst that we can expect from HIV testing as a control strategy.

8.8.1 R_0 and Model Derivation

We first derive an expression for R_0 based on the assumptions in both the Optimistic HIV Test Model and the Pessimistic HIV Test Model. We know from Chapters 3 and 4 that the expected number of addicts infected by a needle in state i infectivity is

$$\frac{\lambda_1 \alpha_i (1-\phi)}{\left(\frac{\tau}{\gamma} + \lambda_1 \hat{\theta}\right)},$$

for $i = 1, 2, 3$ where $\hat{\theta} = 1$ in the Optimistic Model and $\hat{\theta} = \phi$ in the Pessimistic Model. These quantities are the same in our current models, the more complex part of these models comes from the fact that addicts can move permanently from injecting at a rate λ_1 per unit time to a rate λ_2 per unit time at rate δ_t at any time during their infectious lifetime. After an addict is initially infected he or she enters stage one infectivity and at this time will visit shooting galleries and share needles at rate λ_1. As usual consider a newly infected addict entering a population consisting entirely of uninfected addicts and uninfected needles. The addict remains in stage one infectivity injecting at rate λ_1 for on average $1/(\mu + \delta_1 + \delta_t)$ time units, during this period

$$\frac{\lambda_1}{\mu + \delta_1 + \delta_t},$$

needles are infected into stage one infectivity. An addict leaves this state and enters the class of stage one infectivity whose injection rate is λ_2 with probability $\delta_t/(\mu + \delta_1 + \delta_t)$. Once the addict has changed injection rates this is permanent and the addict moves through to stage two and stage three infectivity just as in the Optimistic or Pessimistic Models. Hence the total number of needles of infectivity levels one, two and three which the addict subsequently infects are

$$\frac{\lambda_2}{\mu + \delta_1}, \quad \frac{\lambda_2 \delta_1}{(\mu + \delta_1)(\mu + \delta_2)} \quad \text{and} \quad \frac{\lambda_2 \delta_1 \delta_2}{(\mu + \delta_1)(\mu + \delta_2)(\mu + \delta_3)} \quad \text{respectively.}$$

An addict can also move from stage one infectivity with injection rate λ_1 to stage two infectivity (still with injection rate λ_1), this occurs with probability $\delta_1/(\mu + \delta_1 + \delta_t)$. An

addict remains in stage two infectivity injecting at rate λ_1 for on average $1/(\mu+\delta_2+\delta_t)$ time units and therefore infects

$$\frac{\lambda_1}{\mu+\delta_2+\delta_t},$$

needles while in this class. An addict can then leave this class to join the class where addicts are in stage two infectivity but inject at rate λ_2 with probability $\delta_t/(\mu+\delta_2+\delta_t)$. As previously this move is permanent and hence the total expected number of needles of infectivity levels two and three subsequently infected are

$$\frac{\lambda_2}{\mu+\delta_2} \quad \text{and} \quad \frac{\lambda_2\delta_2}{(\mu+\delta_2)(\mu+\delta_3)} \quad \text{respectively.}$$

Alternatively the addict can remain injecting at rate λ_1 and moves into stage three infectivity with probability $\delta_2/(\mu+\delta_2+\delta_t)$. As previously the addict infects

$$\frac{\lambda_1}{\mu+\delta_3+\delta_t}$$

needles while in this class and moves to injecting at rate λ_2 with probability $\delta_t/(\mu+\delta_3+\delta_t)$. The addict remains in this class until leaving the population and therefore subsequently infects

$$\frac{\lambda_2}{\mu+\delta_3}$$

needles on average into state three infectivity. Therefore the total expected number of previously uninfectious needles left in state one infectivity by an addict during his or her infectious lifetime is

$$\frac{\lambda_1}{\mu+\delta_1+\delta_t}+\frac{\delta_t}{\mu+\delta_1+\delta_t}\frac{\lambda_2}{\mu+\delta_1};$$

the total expected number of previously uninfectious needles left in state two infectivity is

$$\frac{\delta_t\lambda_2\delta_1}{(\mu+\delta_t+\delta_1)(\mu+\delta_1)(\mu+\delta_2)}+\frac{\delta_1\lambda_1}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)}$$

$$+\frac{\delta_1\delta_t\lambda_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)(\mu+\delta_2)};$$

and the total expected number of previously uninfectious needles left in state three infectivity is

$$\frac{\delta_t\lambda_2\delta_1\delta_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}+\frac{\delta_1\delta_t\lambda_2\delta_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)(\mu+\delta_2)(\mu+\delta_3)}$$

$$+\frac{\delta_1\delta_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)}\left[\frac{\lambda_1}{\mu+\delta_3+\delta_t}+\frac{\delta_t\lambda_2}{(\mu+\delta_3+\delta_t)(\mu+\delta_3)}\right].$$

As already mentioned the expected number of addicts infected by each of these three types of infectious needle is the same as in the Optimistic and Pessimistic Models respectively. Therefore

$$
\begin{aligned}
R_0 &= \frac{\lambda_1\alpha_1(1-\phi)}{\left(\frac{\tau}{\gamma}+\lambda_1\hat{\theta}\right)}\left\{\frac{\lambda_1}{\mu+\delta_1+\delta_t}+\frac{\delta_t}{\mu+\delta_1+\delta_t}\frac{\lambda_2}{\mu+\delta_1}\right\} \\
&+ \frac{\lambda_1\alpha_2(1-\phi)}{\left(\frac{\tau}{\gamma}+\lambda_1\hat{\theta}\right)}\left\{\frac{\delta_t\lambda_2\delta_1}{(\mu+\delta_1+\delta_t)(\mu+\delta_1)(\mu+\delta_2)}+\frac{\delta_1\lambda_1}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)}\right. \\
&\qquad \left.+\frac{\delta_1\delta_t\lambda_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)(\mu+\delta_2)}\right\} \\
&+ \frac{\lambda_1\alpha_3(1-\phi)}{\left(\frac{\tau}{\gamma}+\lambda_1\hat{\theta}\right)}\left\{\frac{\delta_t\lambda_2\delta_1\delta_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)}\right. \\
&\qquad +\frac{\delta_1\delta_t\lambda_2\delta_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)(\mu+\delta_2)(\mu+\delta_3)} \\
&\qquad \left.+\frac{\delta_1\delta_2}{(\mu+\delta_1+\delta_t)(\mu+\delta_2+\delta_t)}\left[\frac{\lambda_1}{\mu+\delta_3+\delta_t}+\frac{\delta_t\lambda_2}{(\mu+\delta_3+\delta_t)(\mu+\delta_3)}\right]\right\},
\end{aligned}
$$

with $\hat{\theta}=1$ in the Optimistic HIV Test Model and $\hat{\theta}=\phi$ in the Pessimistic HIV Test Model.

We now state the differential equations which define our three stage HIV testing models. Equations (3.1)-(3.6) can easily be extended to include HIV testing by introducing three additional classes of addicts; one class for addicts who are in stage one infectivity but inject at rate λ_2; one class for addicts who are in stage two infectivity but inject at rate λ_2 and similarly for addicts in stage three infectivity. We denote the fraction of addicts in stage i infectivity who inject at rate λ_1 by π_i^I and those who inject at rate λ_2 by π_i^{II}. Each addict in the population is tested for HIV according to a Poisson process with rate δ_t, therefore addicts in any infectious class move from injecting at a rate λ_1 to injecting at a rate λ_2 with rate δ_t. Therefore the addict equations in the Optimistic and Pessimistic HIV testing models are:

$$
\begin{aligned}
\frac{d\pi_1^I}{dt} &= (1-\pi)\lambda_1(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)(1-\phi)-(\mu+\delta_1+\delta_t)\pi_1^I, \\
\frac{d\pi_2^I}{dt} &= \delta_1\pi_1^I-(\mu+\delta_2+\delta_t)\pi_2^I,
\end{aligned}
$$

$$\frac{d\pi_3^I}{dt} = \delta_2\pi_2^I - (\mu+\delta_3+\delta_t)\pi_3^I,$$

$$\frac{d\pi_1^{II}}{dt} = \delta_t\pi_1^I - (\mu+\delta_1)\pi_1^{II},$$

$$\frac{d\pi_2^{II}}{dt} = \delta_t\pi_2^I + \delta_1\pi_1^{II} - (\mu+\delta_2)\pi_2^{II},$$

and
$$\frac{d\pi_3^{II}}{dt} = \delta_t\pi_3^I + \delta_2\pi_2^{II} - (\mu+\delta_3)\pi_3^{II},$$

where $\pi = \pi_1^I + \pi_2^I + \pi_3^I + \pi_1^{II} + \pi_2^{II} + \pi_3^{II}$. It now remains for us to derive the needle equations for these models according to the different addict-needle interaction assumptions in the Optimistic and Pessimistic Models. Dealing with the Optimistic Model case first, we have the simple assumption that a needle always adopts the infectivity characteristics of the current user. Therefore a needle enters state i infectivity if used by an addict in stage i infectivity and leaves state j infectivity if used by an addict who is not in stage j infectivity, in addition a needle can be exchanged. We now have that addicts in stage i infectivity visit needles at rate $\gamma(\lambda_1\pi_i^I + \lambda_2\pi_i^{II})$ rather than simply $\gamma\lambda\pi_i$. Note also that a needle is used by an uninfectious addict at rate $\lambda_1\gamma(1-\pi)$. Therefore

$$\frac{d\beta_1}{dt} = (\lambda_1\gamma\pi_1^I + \lambda_2\gamma\pi_1^{II})(1-\beta_1) - \beta_1(\lambda_1\gamma\pi_2^I + \lambda_2\gamma\pi_2^{II} + \lambda_1\gamma\pi_3^I + \lambda_2\gamma\pi_3^{II})$$

$$-\gamma\lambda_1\beta_1(1-\pi_1^I-\pi_1^{II}-\pi_2^I-\pi_2^{II}-\pi_3^I-\pi_3^{II}) - \beta_1\tau,$$

$$= \gamma\lambda_1(\pi_1^I-\beta_1) + \lambda_2\pi_1^{II}\gamma + \beta_1\gamma(\lambda_1-\lambda_2)\pi^{II} - \beta_1\tau,$$

where $\pi^{II} = \pi_1^{II} + \pi_2^{II} + \pi_3^{II}$. The other needle equations, $d\beta_2/dt$ and $d\beta_3/dt$ have a similar form. Hence the system of equations which defines the spread of disease where addicts and needles exist in three infectious states, interact according to the assumption of full flushing, and addicts participate in an HIV testing program is:

$$\frac{d\pi_1^I}{dt} = (1-\pi)\lambda_1(\beta_1\alpha_1+\beta_2\alpha_2+\beta_3\alpha_3)(1-\phi) - (\mu+\delta_1+\delta_t)\pi_1^I, \quad (8.45)$$

$$\frac{d\pi_2^I}{dt} = \delta_1\pi_1^I - (\mu+\delta_2+\delta_t)\pi_2^I, \quad (8.46)$$

$$\frac{d\pi_3^I}{dt} = \delta_2\pi_2^I - (\mu+\delta_3+\delta_t)\pi_3^I, \quad (8.47)$$

$$\frac{d\pi_1^{II}}{dt} = \delta_t\pi_1^I - (\mu+\delta_1)\pi_1^{II}, \quad (8.48)$$

$$\frac{d\pi_2^{II}}{dt} = \delta_t\pi_2^I + \delta_1\pi_1^{II} - (\mu+\delta_2)\pi_2^{II}, \tag{8.49}$$

$$\frac{d\pi_3^{II}}{dt} = \delta_t\pi_3^I + \delta_2\pi_2^{II} - (\mu+\delta_3)\pi_3^{II}, \tag{8.50}$$

$$\frac{d\beta_1}{dt} = \gamma\lambda_1(\pi_1^I - \beta_1) + \lambda_2\pi_1^{II}\gamma + \beta_1\gamma(\lambda_1-\lambda_2)\pi^{II} - \beta_1\tau, \tag{8.51}$$

$$\frac{d\beta_2}{dt} = \gamma\lambda_1(\pi_2^I - \beta_2) + \lambda_2\pi_2^{II}\gamma + \beta_2\gamma(\lambda_1-\lambda_2)\pi^{II} - \beta_2\tau, \tag{8.52}$$

and

$$\frac{d\beta_3}{dt} = \gamma\lambda_1(\pi_3^I - \beta_3) + \lambda_2\pi_3^{II}\gamma + \beta_3\gamma(\lambda_1-\lambda_2)\pi^{II} - \beta_3\tau. \tag{8.53}$$

In a similar fashion it is straightforward to derive the system of equations which defines an equivalent model to that above but where addicts and needles interact according to the assumptions in the Pessimistic Model. The needle equations are similar to those in (4.4)-(4.6) except that as above we now have that addicts in stage i visit needles at rate $\gamma(\lambda_1\pi_i^I + \lambda_2\pi_i^{II})$ and a needle is used by an uninfectious addict at rate $\lambda_1\gamma(1-\pi)$. Needles enter state one infectivity after use by an addict in stage one infectivity. Needles leave state one infectivity after use by an addict who is in stage two infectivity, stage three infectivity or is uninfectious and cleans the needle prior to use, in addition needles may be exchanged. Hence

$$\begin{aligned}\frac{d\beta_1}{dt} &= \gamma(1-\beta_1)(\lambda_1\pi_1^I + \lambda_2\pi_1^{II}) - \beta_1(\lambda_1\pi_2^I + \lambda_2\pi_2^{II})\phi\gamma - \beta_1(\lambda_1\pi_3^I + \lambda_2\pi_3^{II})\phi\gamma \\ &\quad -\beta_1(1-\pi)\lambda_1\phi\gamma - \beta_1\tau.\end{aligned}$$

The equations for needles in state two and three infectivity follow in a similar manner by replacing the terms in eqns (4.5)-(4.6) with the new visiting rates. Therefore the system of equations which defines the spread of disease where addicts and needles exist in three infectious states, interact according to the assumptions in the Pessimistic Model, and addicts participate in an HIV testing program is:

$$\frac{d\pi_1^I}{dt} = (1-\pi)\lambda_1(\beta_1\alpha_1 + \beta_2\alpha_2 + \beta_3\alpha_3)(1-\phi) - (\mu+\delta_1+\delta_t)\pi_1^I, \tag{8.54}$$

$$\frac{d\pi_2^I}{dt} = \delta_1\pi_1^I - (\mu+\delta_2+\delta_t)\pi_2^I, \tag{8.55}$$

$$\frac{d\pi_3^I}{dt} = \delta_2\pi_2^I - (\mu+\delta_3+\delta_t)\pi_3^I, \tag{8.56}$$

$$\frac{d\pi_1^{II}}{dt} = \delta_t\pi_1^I - (\mu+\delta_1)\pi_1^{II}, \tag{8.57}$$

$$\frac{d\pi_2^{II}}{dt} = \delta_t\pi_2^I + \delta_1\pi_1^{II} - (\mu+\delta_2)\pi_2^{II}, \tag{8.58}$$

$$\frac{d\pi_3^{II}}{dt} = \delta_t\pi_3^I + \delta_2\pi_2^{II} - (\mu+\delta_3)\pi_3^{II}, \tag{8.59}$$

$$\begin{aligned}\frac{d\beta_1}{dt} &= (1-\beta_1)(\lambda_1\pi_1^I + \lambda_2\pi_1^{II})\gamma - \beta_1(\lambda_1\pi_2^I + \lambda_2\pi_2^{II})\phi\gamma - \beta_1(\lambda_1\pi_3^I + \lambda_2\pi_3^{II})\phi\gamma \\ &\quad -\beta_1(1-\pi)\lambda_1\phi\gamma - \beta_1\tau,\end{aligned} \tag{8.60}$$

$$\begin{aligned}\frac{d\beta_2}{dt} &= (1-\beta_1-\beta_2-\beta_3)(\lambda_1\pi_2^I + \lambda_2\pi_2^{II})\gamma + \beta_1(\lambda_1\pi_2^I + \lambda_2\pi_2^{II})\phi\gamma \\ &\quad +\beta_3(\lambda_1\pi_2^I + \lambda_2\pi_2^{II})\phi\gamma - \beta_2(\lambda_1\pi_3^I + \lambda_2\pi_3^{II})\gamma - \beta_2(\lambda_1\pi_1^I + \lambda_2\pi_1^{II})\gamma \\ &\quad -\beta_2(1-\pi)\lambda_1\gamma\phi - \beta_2\tau,\end{aligned} \tag{8.61}$$

and

$$\begin{aligned}\frac{d\beta_3}{dt} &= (1-\beta_1-\beta_3)(\lambda_1\pi_3^I + \lambda_2\pi_3^{II})\gamma + \beta_1(\lambda_1\pi_3^I + \lambda_2\pi_3^{II})\gamma\phi - \gamma\beta_3(\lambda_1\pi_1^I + \lambda_2\pi_1^{II}) \\ &\quad -\beta_3(\lambda_1\pi_2^I + \lambda_2\pi_2^{II})\gamma\phi - \beta_3(1-\pi)\gamma\phi\lambda_1 - \beta_3\tau.\end{aligned} \tag{8.62}$$

The systems of equations in (8.45)-(8.53) and (8.54)-(8.62) are sufficiently complex that it seems appropriate to limit our study of these models to the use of numerical integration. We are interested in two main aspects of these models, the first and more important of these is the effect that different relative infectivity assumptions have on the long term prevalence of disease when addicts are regularly tested for HIV. Secondly and to a lesser extent we are are also interested in whether $R_0 = 1$ is still the critical threshold criteria between the disease-free and endemic states in our models. Given that this was the case in the Optimistic, Pessimistic and HIV Test Models it seems highly likely that this will be the case here also.

8.8.2 Simulations

We now simulate the Optimistic and Pessimistic HIV testing models using $\lambda_1 = 250$ per year, $\lambda_2 = 150$ per year and where addicts are tested on average once every ten years ($\delta_t = 0.1$). The other parameter estimates are as in Table 3.1 in Chapter 3. Figures 8.9 and 8.10 show simulations of the total fraction of infected addicts and the total fraction of infected needles in the Optimistic and Pessimistic HIV Test Models respectively. In each figure it was assumed that initially 1% of all addicts were infectious and all these

Figure 8.9: Three Stage Infectivity and HIV Testing: Optimistic ($R_0 > 1$)

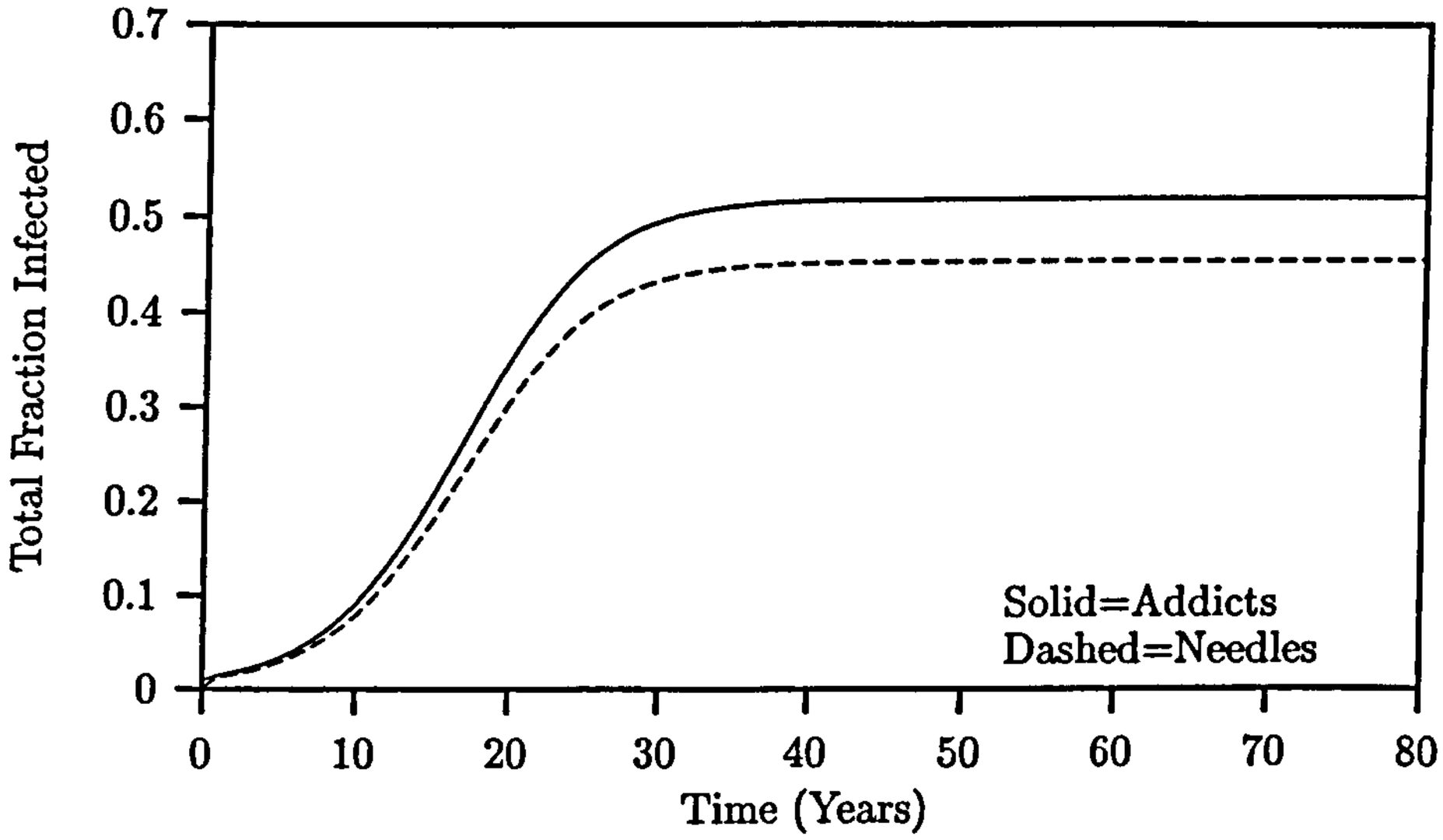
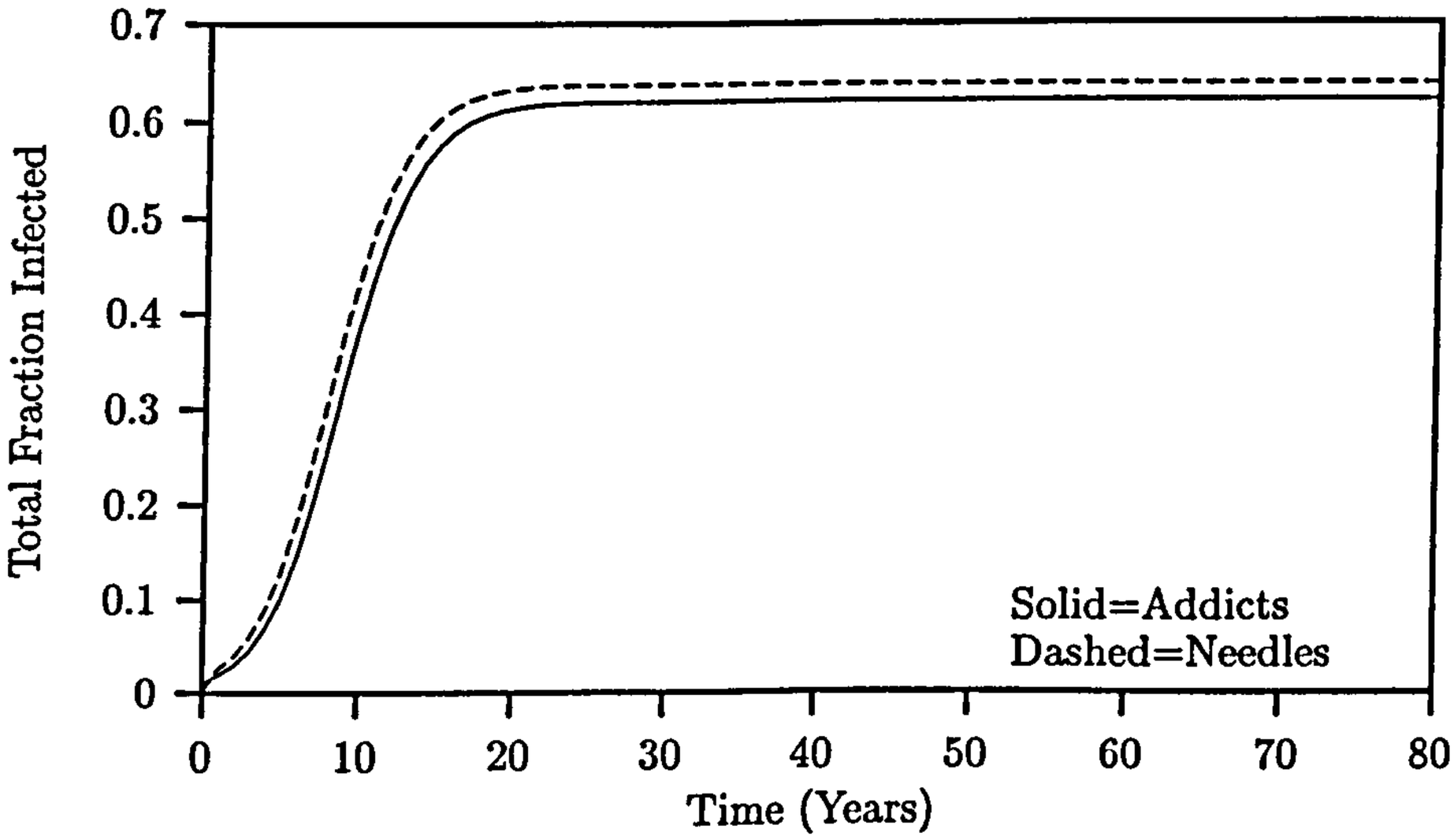


Figure 8.10: Three Stage Infectivity and HIV Testing: Pessimistic ($R_0 > 1$)



addicts were in stage one infectivity and were unaware of their HIV positive status. These simulations suggest that as in all our previous models the prevalence of disease in both addicts and needles eventually reaches an endemic steady state. The values for R_0 in each of these models using the current set of parameter estimates are 1.95 and 2.94 for the Optimistic and Pessimistic HIV Test Models respectively and the steady state solutions are $(\pi^\star, \beta^\star) = (0.517, 0.453)$ and $(\pi^\star, \beta^\star) = (0.619, 0.636)$ respectively.

We do not illustrate any more simulations of the Optimistic and Pessimistic HIV Test Models, however many simulations of these models were carried out using a variety of different parameter estimates and initial conditions. In each case we found that if $R_0 > 1$ (and disease is initially present) then the disease tends to an endemic steady state and if $R_0 < 1$ then the disease eventually dies out in all addicts and all needles. Having demonstrated that our two models exhibit similar long term behaviour to all our previous models we now focus on the main objective of this section, namely determining what effect different relative infectivity assumptions have on the long term prevalence of disease when addicts are tested regularly for HIV.

8.8.3 Relative Infectivity Assumptions

We now examine the long term prevalence of HIV in our two models using a variety of different relative infectivity assumptions. The relative infectivity of an addict in each of the three infectious stages is denoted by the HIV transmission probabilities α_1, α_2 and α_3 respectively. By adjusting the values of these three parameters we can vary the relative infectivity of addicts in each of the three infectious stages. If we are to use different infectivity assumptions then the values of α_1, α_2 and α_3 require some form of standardisation in order to produce fair comparisons.

In order to estimate the HIV transmission probability α in Kaplan and O'Keefe (1993) a model based estimation technique was used. This method assumed that the model under study was a reasonable approximation to reality and therefore the endemic equilibrium solution of this model should be close to the observed prevalence of infectious addicts and needles (assuming of course that the population under study is at equilibrium). Hence if all other model parameters have been estimated then α can be estimated by solving for α in the equation $\beta^\star = \beta^\dagger$ (where $\beta^\star$ is a known function of the model parameters and $\beta^\dagger$ the observed prevalence of disease in needles). This method was used by Kaplan and O'Keefe (1993) who were able to estimate the prevalence of

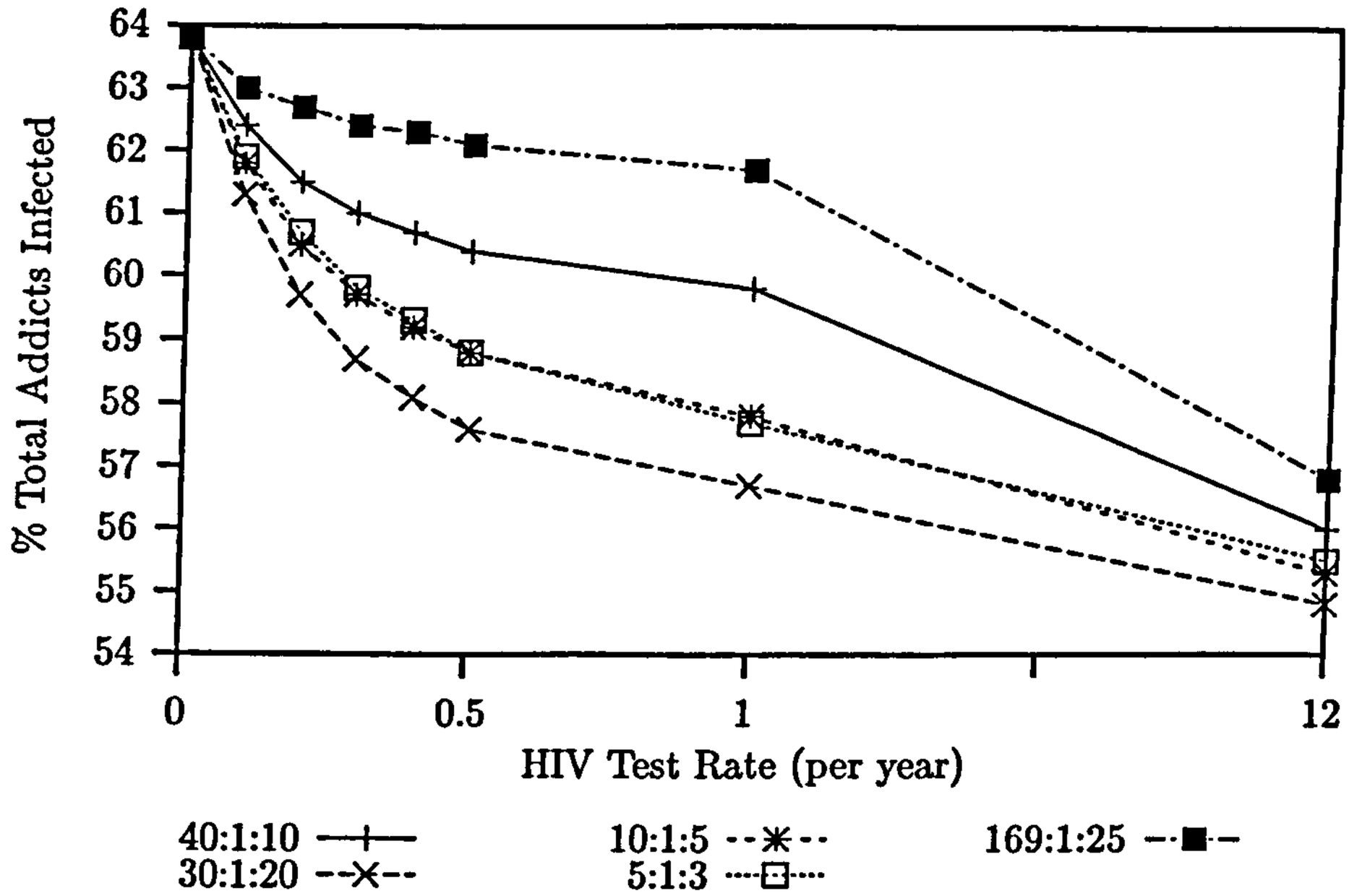
Table 8.2: Relative Infectivity Ratios and HIV Transmission Estimates

α_1:α_2:α_3	α_1	α_2	α_3
40: 1 :10	0.03300	0.00083	0.00825
30: 1 :20	0.01610	0.00054	0.01072
10: 1 :5	0.01813	0.00181	0.00906
5: 1 :3	0.01412	0.00282	0.00847
169: 1 :25	0.04743	0.00028	0.00702

infectious needles circulating among the addict population using data from the New Haven needle exchange program. An obvious problem with this method is that the estimated value of α will vary depending on the model used, however in the absence of a better alternative this method seems appropriate. We used a similar method to estimate α_2 in our three stage models (see Appendix B for details). We estimated all other model parameters (including ζ_1 and ζ_3 where $\alpha_1 = \zeta_1\alpha_2$ and $\alpha_3 = \zeta_3\alpha_2$) and solved to find an estimate of α_2 which gave rise to $\beta^\star = \beta^\dagger$ (using the Pessimistic Model from Chapter 4). Kaplan and O'Keefe (1993) estimate that $\beta^\dagger = 0.675$ for the drug using population in New Haven, Connecticut, USA, prior to the introduction of a formal needle exchange program. Using this estimation method for various different choices of ζ_1 and ζ_3 will provide us with estimates of α_1, α_2 and α_3 which are standardised. The standardisation is that each set of values of α_1, α_2 and α_3 gives rise to an equilibrium prevalence of $\beta^\star = 0.675$ in the Pessimistic Model (without HIV testing but including needle cleaning and a natural turnover rate of needles).

We now examine the prevalence of HIV in addicts in our two models using five different relative infectivity assumptions. These assumptions are shown in Table 8.2. We always assume that stage two infectivity is the least infectious followed by stage three then stage one. Our five infectivity ratios have been chosen to provide a varied range of assumptions. The penultimate infectivity assumption of $5 : 1 : 3$ was the one used by Peterson et al. (1990) and the final infectivity assumption is similar to that suggested by Seitz and Müller (1994). However note that while Seitz and Müller use a three stage infectious period in their model their third stage of infectivity refers to full blown AIDS rather than Pre-AIDS symptoms. We additionally assume that on average an addict spends 2.6 months in stage one, 52.6 months in stage two and 62.5 months in

Figure 8.11: Pessimistic Assumptions: $\lambda_2 = 150$ per year

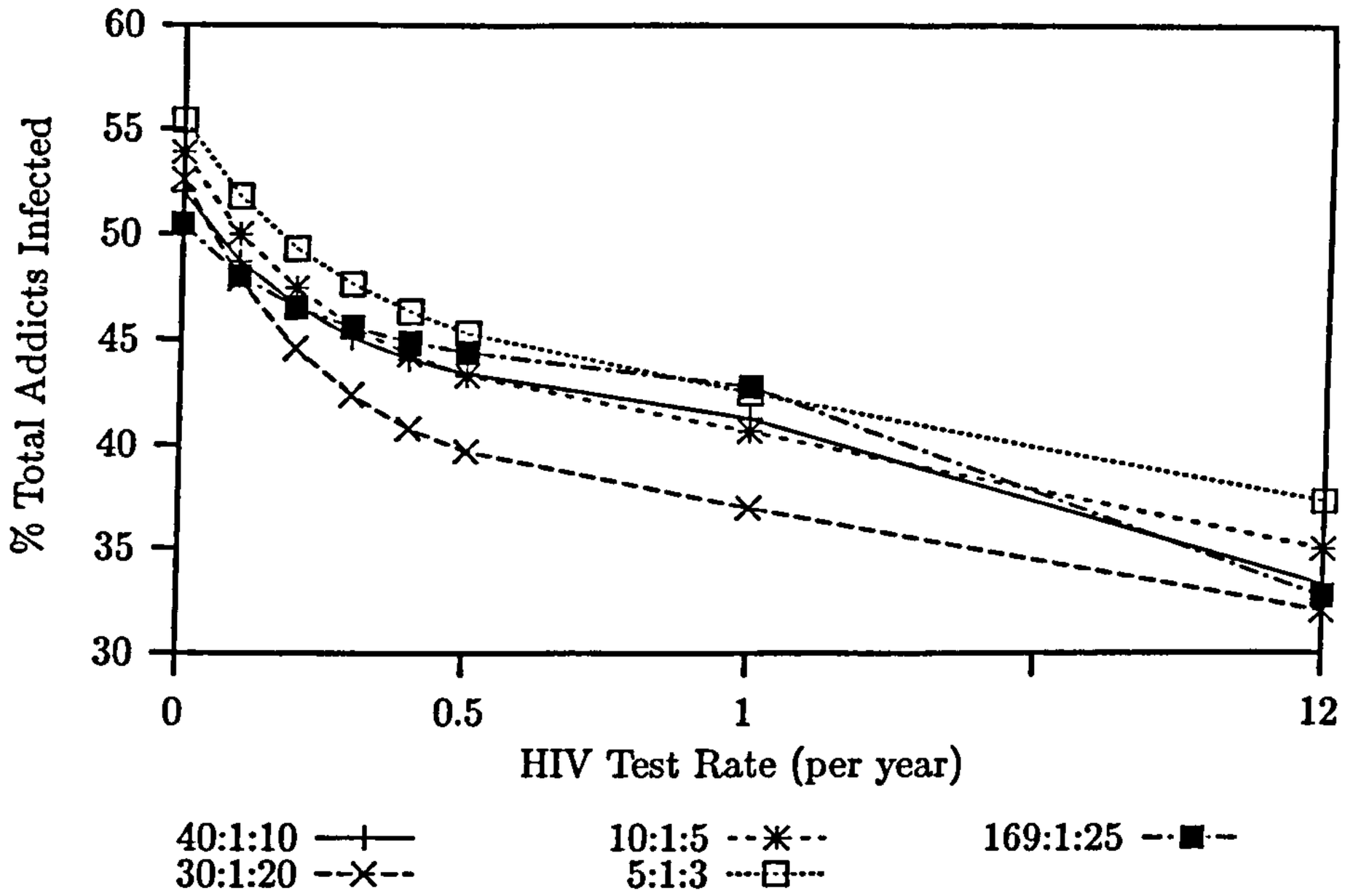


stage three (as estimated by Peterson et al. (1990) and used in our previous models).

Figures 8.11 and 8.12 show the total long term prevalence of HIV among addicts for the five infectivity assumptions detailed in Table 8.2 and for addict HIV test rates of 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0 and 12.0 per year. In each case we assume that addicts who have not yet tested positive for HIV have a shared injection rate of $\lambda_1 = 250$ per year and those who are aware of their HIV positive status inject with a shared needle on average $\lambda_2 = 150$ times per year. All other model parameters are as detailed in Table 3.1 in Chapter 3. Figure 8.11 shows the long term prevalence in the Pessimistic HIV Test Model and Figure 8.12 is equivalent but uses the Optimistic HIV Test Model.

The first point of interest in these figures is that the relative infectivity ratio of 30:1:20 appears to be the most affected by HIV testing in both models and at all testing rates. This is not surprising as this assumption is the most heavily loaded towards the end of the AIDS incubation period. Therefore HIV testing has the most to offer by removing addicts prior to entering stage three infectivity. Perhaps more surprising is that infectivity assumptions which have a higher prevalence in the pessimistic model do not necessarily have a higher prevalence in the optimistic model. Furthermore there appears to be a large difference in the effectiveness of HIV testing between these two models. This suggests that addict-needle interaction assumptions may have a significant influence on the endemic equilibrium prevalence of disease irrespective of any particular

Figure 8.12: Optimistic Assumptions: $\lambda_2 = 150$ per year



relative infectivity assumptions.

Perhaps the most interesting feature in Figures 8.11 and 8.12 is the rather modest effect which HIV testing has in each of the five infectivity assumptions. In particular the reduction in disease in the Pessimistic HIV Test Model is very small even when addicts are tested as frequently as monthly. The Optimistic HIV Test Model is more responsive but even in this case the relative reduction in the long term prevalence of disease is not great. In summary, it seems that for the current values of λ_1 and λ_2 and our five relative infectivity assumptions, the testing of addicts for HIV will not produce a substantial decrease in the level of disease.

In Section 8.7 we argued that both the HIV test rate and the post-test shared injection rate must be low for this control strategy to provide a significant decrease in the level of disease. Therefore we now examine the long term prevalence in HIV in our two models in a similar manner to previously but now we assume that once aware of their infectious status addicts only inject about $\lambda_2 = 50$ times a year on average. Figures 8.13 and 8.14 were constructed in a similar fashion to Figures 8.11 and 8.12 but where $\lambda_2 = 50$ per year. The most obvious difference between the previous figures and Figures 8.13 and 8.14 is that now when the HIV test rate is increased there is a large drop in the prevalence of disease, in particular the disease dies out for each infectivity assumption in the Pessimistic HIV Test Model when addicts are tested on average once

Figure 8.13: Pessimistic Assumptions: $\lambda_2 = 50$ per year

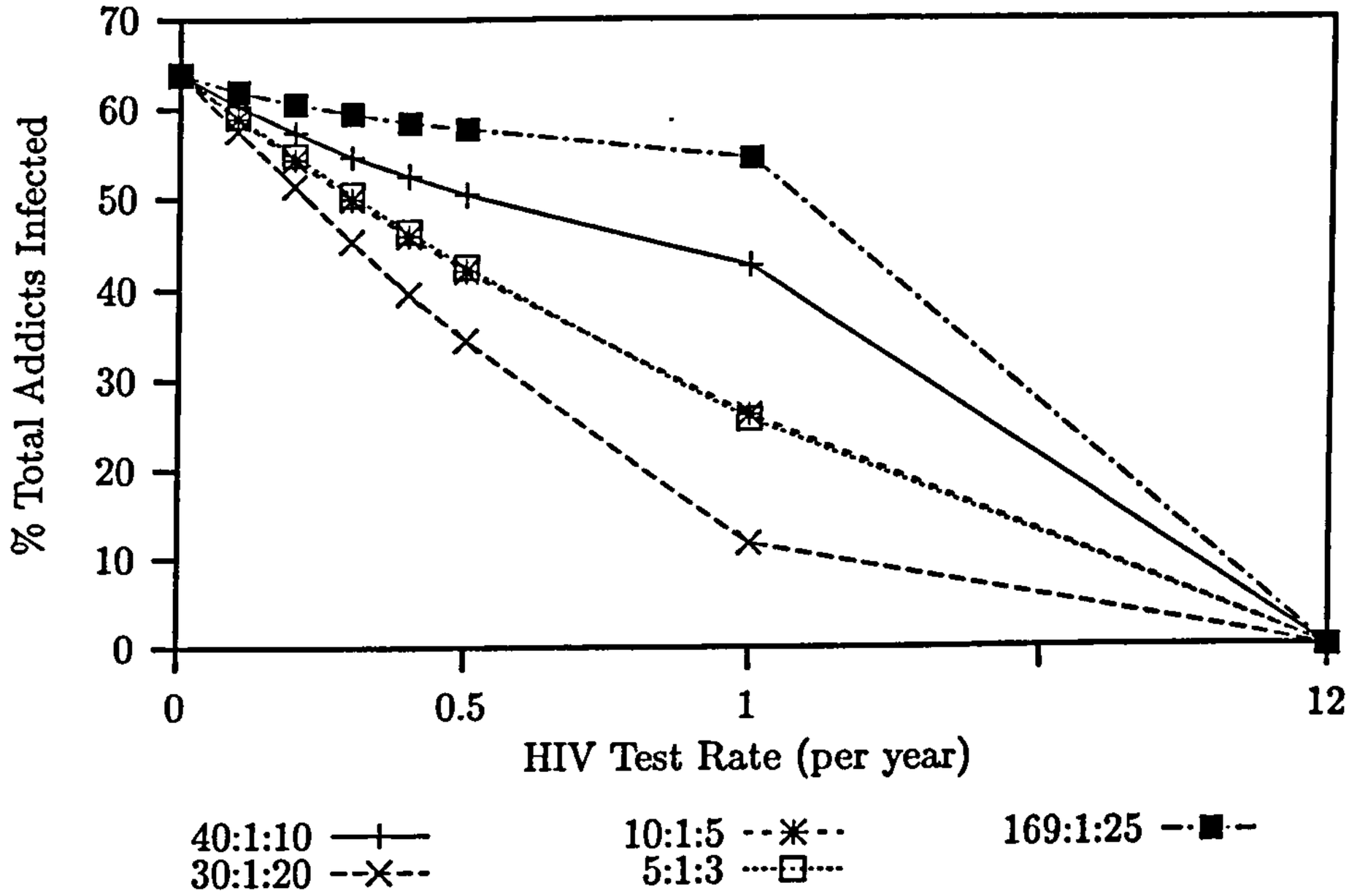
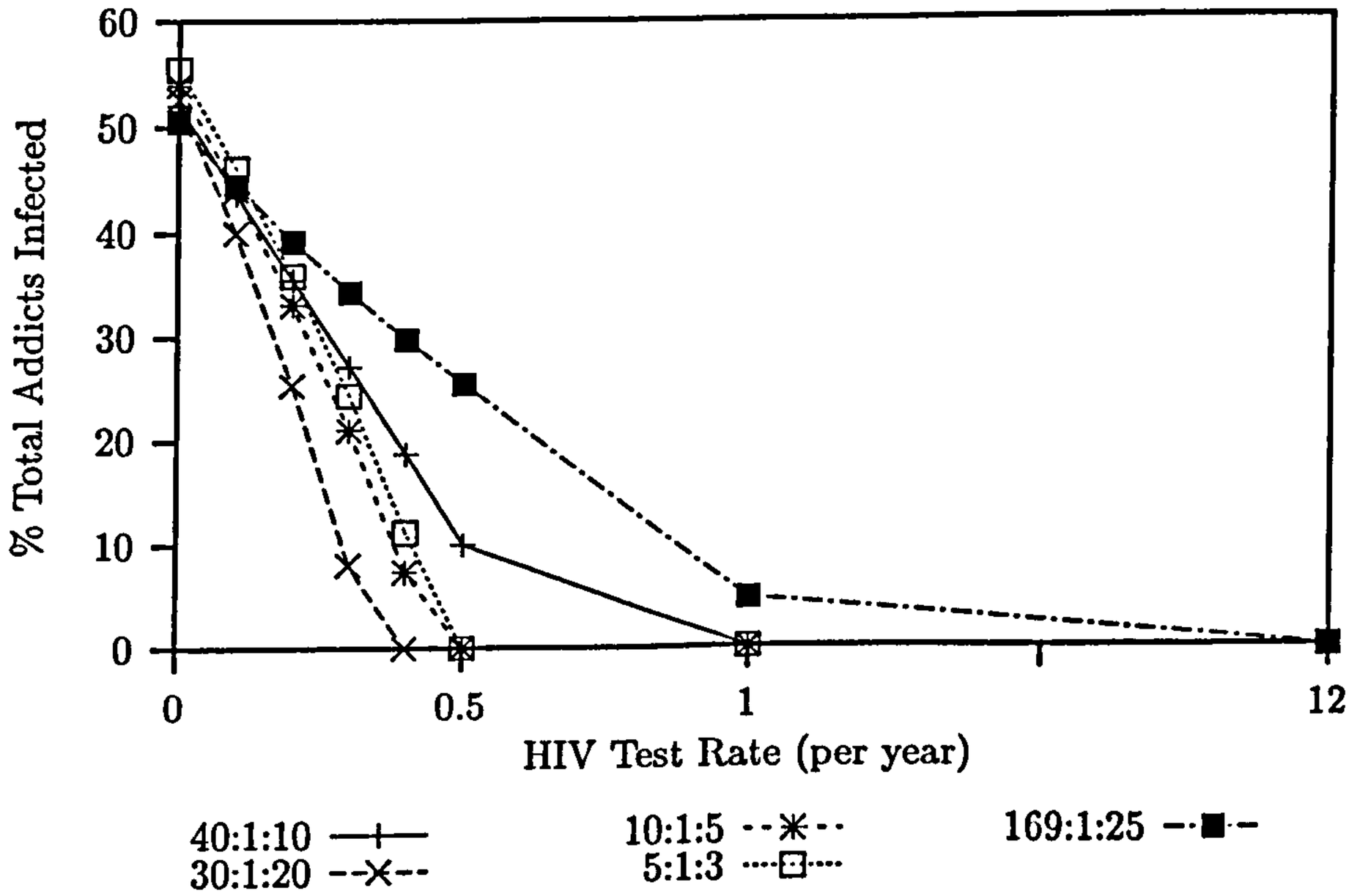


Figure 8.14: Optimistic Assumptions: $\lambda_2 = 50$ per year



a month. In the Optimistic HIV Test Model the disease can die out at an HIV test rate of only 0.4 per year. These figures again illustrate that for HIV testing to be an effective control strategy we require that addicts are tested regularly and moreover that once tested the shared injection rate must be very substantially reduced. In Figures 8.13 and 8.14 it is now also much clearer as to which infectivity assumptions are the most affected by HIV testing. As previously we have that the infectivity ratio 30:1:20 is the most affected, the figures now also suggest that the ratios next most affected are 10:1:5 followed by 5:1:3 then 40:1:10 and finally 169:1:25. This order seems sensible since 10:1:5 and 5:1:3 have proportionally more infectivity in stage three than in 40:1:10 which in turn has more infectivity in stage three than 169:1:25.

8.9 Summary of Results for the HIV Testing Models

We first briefly discussed a model by Greenhalgh and Hay (1997) which examined the effect of HIV testing, but which assumed that the proportion of infectious addicts who have tested positive for HIV is constant throughout the duration of an HIV epidemic. We extended this model to allow for addicts to be tested for HIV at a constant rate per unit time, where a positive test results in an addict changing the rate at which he or she shares needles. Incorporating HIV testing in this fashion allows the proportion of infectious addicts who have tested positive for HIV to vary throughout the course of an HIV epidemic which is the more realistic situation. After deriving this extended model we then examined its long term behaviour and computed an expression for the basic reproductive number. We showed that if $R_0 \leq 1$ then disease will eventually die out in all addicts and all needles. If $R_0 > 1$ then the endemic equilibrium is locally stable to small perturbations, and moreover if disease is initially present then it will persist indefinitely among the population. We then used simulation to confirm that if $R_0 \leq 1$ then the disease-free equilibrium is globally stable and examined the long term behaviour of the model when $R_0 > 1$. Simulations suggest that when $R_0 > 1$ the prevalence of disease always tends to the endemic equilibrium solution.

We next compared the behaviour of our extended model with the original model due to Greenhalgh and Hay. By using a suitable calibration method we showed that the long term prevalence of disease in both models is identical, however simulations suggest that the disease spreads more slowly in the original model. We then demonstrated that for HIV testing to be an effective control strategy in our model it is necessary that

addicts are both tested frequently for HIV, and once aware of their infectious status they substantially reduce the rate at which they inject with shared needles.

We finished off this chapter with a brief look at combining the three stage infectivity models discussed in Chapters 3 and 4 with HIV testing. We stated the differential equations which defined these two models and derived an expression for the basic reproductive number. Simulations suggest that again $R_0 = 1$ is the critical threshold between the disease-free and endemic states. Simulations also suggest that as with our previous models if $R_0 > 1$ then the prevalence of disease tends to the endemic equilibrium solution. We investigated the effect that different relative infectivity assumptions have on the long term prevalence of disease when addicts are tested regularly for HIV. We found three main points of interest. Firstly for HIV testing to be an effective control strategy we require that addicts are tested frequently and once aware of their infectious status addicts must share needles very infrequently. Secondly it appears that addict-needle interaction assumptions can have a significant impact on the effectiveness of HIV testing. For example HIV testing appears to be very much more effective when addicts and needles interact as in the Optimistic Model compared to the interaction assumptions in the Pessimistic Model. Thirdly and finally, the choice of relative infectivity assumptions also has a significant impact on the effectiveness of HIV testing. We find that those infectivity assumptions which result in addicts having a relatively high viral load immediately prior to developing full blown AIDS result in the largest benefit from HIV testing.

Chapter 9

Sensitivity Analysis of Deterministic Models

9.1 Introduction

The aim of studying models of HIV transmission is to provide information which can assist public health officials and others to develop policies for disease control and highlight beneficial areas of future research. We have spent considerable effort in the previous chapters establishing that the spread of HIV among drug users is governed by a threshold effect (threshold $R_0 = 1$) when addicts are allowed to progress through three stages of infectivity. The most immediate practical use of this information is that it provides a goal for any control strategy, in that reducing R_0 to less than unity will (eventually) eradicate HIV. While this information is useful and adds to the numerous other examples of epidemiological models which have $R_0 = 1$ as a threshold condition, we would also like to use our models to extract more specific information. For example, it would be useful for us to be able to determine which model parameters are the most influential in affecting the spread of HIV. This information could then be used to focus control strategies on areas which can deliver the most benefit. Ideally we should also like to be able to use our HIV transmission models to assist in predicting the future number of cases of HIV and AIDS. Unfortunately, as has been noted by several authors (Blower et al., 1991, Anderson and May, 1991) in order for HIV transmission models to provide useful estimates of the future number of HIV and AIDS cases very accurate data are required, and as yet such data are not available.

In this chapter we discuss a method proposed by Blower and Dowlatabadi (1994)

to determine which model parameters are the most influential in affecting the spread of HIV, and use the HIV Test Model as an illustration. Ideally we should like to carry out a thorough sensitivity analysis on all the models we have discussed in this thesis, however due to a lack of data, specifically data relating to the sampling distributions of the parameters in our previous models, we limit our study to examining the methodology proposed by Blower and Dowlatabadi. What is of particular interest is that from an experimental design perspective this method requires some care in order that meaningful results are obtained.

We first discuss an analogy between experimental trials and the sensitivity analysis of epidemiological models. We then give some background information on several common types of experimental design and discuss the differences between these designs. We next outline the Latin Hypercube sampling scheme proposed by Blower and Dowlatabadi (1994) and we discuss the appropriateness of this design. We then conduct a limited sensitivity analysis on the HIV Test Model to illustrate the potential problems of using Latin Hypercube sampling to determine which model parameters are the most influential in affecting the spread of HIV. We conclude with a discussion on how the method by Blower and Dowlatabadi can still be used to good effect if several factors are taken into consideration.

9.2 Experimental Design Analogy

Suppose that we have a complex HIV transmission model with many parameters. Suppose also that surveys have been conducted so that we have preliminary information about each parameter in our model and moreover we have a rough estimate of the sampling distribution of each parameter. For example suppose that we have surveyed a number of different groups of addicts and have collected estimates of the mean shared injection rate in each of these different groups. From this we have an estimate of the sampling distribution of λ and a point estimate of λ (the mean of the sampling distribution, say). We now wish to simulate our model to determine which parameters are the most influential in affecting the future number of cases of HIV. This situation is very similar to that of an agricultural researcher (say) with a field of crops who wishes to determine which fertiliser (or combination of fertilisers) causes the greatest increase in crop yield. In this analogy the crop yield represents the future prevalence of HIV, the different types of fertiliser the different parameters in our model, and the

size of the field represents the number of simulations of our model. This agricultural analogy has been studied very extensively and the design and analysis of experiments to determine which treatments (or fertilisers) are the most effective represents a large body of work pioneered by R. A. Fisher at the Rothamstead Institute (see Fisher and Bennett, 1990). Hence it seems natural to use some of this research to assist in devising a suitable method from which to assess the importance of each of the parameters in a model of HIV transmission.

Whilst the aim of a sensitivity analysis and an agricultural trial are very similar there are several important differences in the framework of these two approaches. Firstly consider an agricultural trial. The researcher has a field of crops and he or she decides which treatments or combinations of treatments to apply to each area of the crop. We can consider the crop yield in any particular part of the field as a function of the treatments which have been applied. It is also appropriate to assume that the response (the crop yield) from any particular combination of treatments is a random variable. Now consider a sensitivity analysis. We have that each input parameter follows a particular sampling distribution. If we are using a deterministic HIV transmission model then the response variable (the future number of HIV cases, say) can be thought of as simply a deterministic mapping from the k parameter estimates in the model to a single number. In each case we wish to determine which treatments (parameters) have the largest effect on the response variable. The main difference between these two analogies is that in an agricultural trial the inputs are fixed by the experimenter and the output is a random response of these fixed inputs. In contrast, in a sensitivity analysis the inputs are not fixed but are chosen according to the sampling distributions of the various model parameters and the response is a deterministic function of the inputs. Hence it could be argued that the theoretical framework for a sensitivity analysis is actually back-to-front from that of a conventional experimental trial. Blower and Dowlatabadi (1994) argue that a sensitivity analysis can be considered in a similar statistical framework to experimental trials and it is on this basis that we now proceed.

9.3 Experimental Design Choice

We now discuss several elementary experimental designs and how these designs can be used to deliver information relating to the relative importance of different factors.

Experimental designs can be split into two broad categories, those concerned with one factor of interest and those concerned with multiple factors of interest. Consider the agricultural analogy described above. Suppose that we want to investigate whether different concentrations of a particular fertiliser affect the crop yield. In this situation we could use an experimental design which has only one factor of interest with a number of different treatment levels (our one factor being the single type of fertiliser used). Using an equivalent sensitivity analysis analogy we have a single parameter which we divide into distinct subgroups. We then investigate whether the difference in the crop yield (or equivalently the long term prevalence of HIV) between each of the treatment levels is statistically significant. These are examples of experiments where a single factor is of interest. Suppose now that we have two different fertilisers each with four concentration (treatment) levels and we wish to determine which fertiliser or combination of fertiliser causes the greatest increase in crop yield. In this situation we should employ a multiple factor of interest experimental design. An equivalent sensitivity analysis analogy is that we now have two parameters each of which has been split into four subgroups. We now discuss in some detail designs for single and multiple factor experiments. This discussion leads us on to describing the experimental method advocated by Blower and Dowlatabadi and the potential problems of using this method.

Single Factor and Multiple Factor Experiments

Each experimental design corresponds to the fitting of a general linear model (g.l.m.) to the response variable. The fundamental difference between a single factor and multiple factor experiment is in the form of g.l.m. that we use to model the data. The simplest form of g.l.m. is the following:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \tag{9.1}$$

where y_{ij} represents the j^{th} observation of the response variable at treatment level i, μ is the overall mean of the observed data, α_i is the (mean) effect from treatment level i, and ϵ_{ij} is a random error term. This g.l.m. corresponds to a number of scenarios, for example we might want to investigate whether altering the concentration of a particular fertiliser treatment has a significant effect on the crop yield and if so which concentration is the most effective. Hence we wish to establish whether there is a statistically significant difference in the response variable between treatments levels (concentrations of fertiliser). This situation represents an experiment where only one factor is of inter-

est, that of the effect on the crop yield of the concentration of fertiliser being applied. We investigate this effect by determining whether any α_i terms for $i = 1, 2, 3, 4$ (using four treatment levels) in the above g.l.m. are statistically significant. This is achieved by examining the variation in the observed data between treatment levels compared to the variation within treatment levels. This is a standard method referred to as analysis of variance (ANOVA). If the variation between treatment levels is large enough compared to the variation within each treatment level then we have a statistically significant treatment effect, in other words we find that varying the concentration of fertiliser does have a statistically significant effect on crop yield.

Suppose that we carried out the above experiment and found that the treatment effect was not statistically significant. The reason for this could be that the variation in the observed data within each treatment level (the residual variation in the data) was so large that an effect due to altering the treatment level could not be detected. Our previous design assumed that our field of crops was homogeneous, however this is unlikely in real life. For example some parts of our field may have access to a water source such as an underground spring, while some parts may be protected from the wind by a shelter belt of trees. It is reasonable to assume that these two extra factors may have caused a substantial increase in the variability of the crop yield throughout our field and as such reduced the power of our previous statistical test. We could express this as "not being able to see the wood for the trees".

The three factors in our experimental trial are now the fertiliser treatment, the water source and the shelter belt. We can use a very common experimental design, the Latin Square design, to effectively investigate this more complex situation. By using this design we are assuming that

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \tag{9.2}$$

where y_{ijk} represents the yield from using treatment i in row j and column k of the (square) field, μ is the overall mean yield, α_i is the (mean) effect from treatment i, β_j is the (mean) effect from being in row j, γ_k is the (mean) effect from being in column k and finally ϵ_{ijk} is a random error term. It is important to note carefully the form of this model, it assumes that the effect on crop yield due to access to the water source (the row effect, β_j) is the same irrespective of how close to the shelter belt the crop resides. Similarly the effect on crop yield from being sheltered from the wind (the column effect, γ_k) is the same irrespective of how close the crop is to the water source,

and finally the effect due to the fertiliser treatment level (α_i) is the same irrespective of the location of the crop to either the shelter belt or the water source. Hence this design and corresponding g.l.m. assumes that each of the three effects (or factors) on the response variable act independently of each other, we have no interaction effect between the various factors in our model.

In our example we have introduced the water source and shelter factors purely to reduce the residual variation in our observed data and improve the assessment of our single factor of interest, the effect due to changing treatment levels. In this experiment the factors, water source and shelter, are referred to as blocking factors, factors introduced purely to reduce the residual (background) variation in the observed data and thus give a better fitting model, which allows for a more accurate assessment of the factor of interest. In effect the total variation in the data can be broken down into the variation due to the water source, shelter and the fertiliser treatment. By removing the former two sources of variation we can greatly improve the accuracy of our statistical test for the effect due to changing treatment levels.

To summarise, a single factor experiment represents the situation where the response variable is simply the sum of independent effects (plus a random error term). We have one main factor of interest and the remaining factors are present to remove residual variation in the data and improve the fit of the g.l.m. We now discuss the more complex situation of a multiple factor experiment. In this case we no longer assume that the effects of the experimental factors act independently but that interaction may occur between these factors. Hence again using the agricultural analogy, we now allow for the effect due to treatment level i to depend also on how much shelter the crop receives and how much water the crop has access to.

In deciding whether to conduct a single or multiple factor experiment it is first necessary to establish whether it is reasonable for an interaction affect to occur between the various experimental factors. In the above agricultural example it is probably quite satisfactory to assume that the factors act independently and use a single factor (Latin Square) design. However now suppose that the researcher has two types of fertiliser which can each be applied at four concentrations, and wishes to determine which combinations of fertiliser are the most promising for further research. In this experiment it is wiser to at least initially assume that the effect due to each fertiliser may not act independently. A standard way of conducting this experiment would

be to use a 2^4 factorial design. This design involves applying each fertiliser at four concentrations and each part of the field is treated with a different combination of fertilisers so that all sixteen permutations have been tried. In addition we treat at least two plots with each treatment combination. In terms of a model this design assumes the following:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \tag{9.3}$$

where y_{ijk} represents the response in factor one level i and factor two level j and k denotes the observation number in cell (i, j) (in our example each cell has at least two observations). As above we have that μ is the overall mean response, α_i is the effect from factor one at level i, β_j is the effect from factor two at level j, and ϵ_{ijk} is a random error term. The $(\alpha\beta)_{ij}$ term represents the interaction effect between factor one and factor two at level i and j respectively. Since we tested at least two plots with each treatment combination we can use ANOVA in a similar manner as in the Latin Square design to test for the statistical significance of each of the terms in our g.l.m. In particular we first test to see whether the interaction term is significant, if this term is not significant then we could if desired repeat the experiment with a single factor of interest design and not lose any accuracy, as the data suggests that the factors do act independently. The inclusion of an interaction factor is the fundamental difference between a single factor and multiple factor design.

We have illustrated several experimental designs and the general linear models which correspond to these designs. It should be clear that how we expect the experimental factors to affect the response variable governs the choice of whether to use a simpler single factor design or a more complex multiple factor design. In other words if we strongly believe that the experimental factors will act independently of each other then a single factor design should be used, if this is not the case then it is more appropriate to use a multiple factor design. The main practical disadvantage of using multiple factor designs (for example factorial designs) is that a large number of observations are required in order that the interaction terms in the g.l.m. can be tested. A common method of dealing with experiments with many different factors is to initially use a full factorial scheme with each factor limited to only two levels. The resulting observations are then studied and the factors or combinations of factors which are not statistically significant are then dropped from the g.l.m. and a more detailed experiment is carried out using only the factors which have been shown to be more important.

The aim of a sensitivity analysis is to determine which parameters are most influential in affecting the outcome variable. A fundamental question in the context of this chapter is whether we expect the model parameters in complex models of disease transmission to each act independently on the outcome variable. It seems reasonable to argue that this will *not* be the case. It seems more likely that the level of disease in the population at any time will be a possibly very complex non-linear function of the model parameters, and as such the effect of increasing or decreasing any one model parameter will also depend heavily on the current values of other model parameters. Hence it would seem appropriate that a sensitivity analysis should be conducted using a multiple factor experimental design. We now discuss in detail the experimental design proposed by Blower and Dowlatabadi (1994).

9.4 Latin Hypercube Design

Blower and Dowlatabadi propose the use of a Latin Hypercube (LHC) design together with partial rank correlation coefficients (PRCC) to determine the most influential model parameters in complex HIV transmission models. A Latin Hypercube design is an extreme version of the Latin Square design discussed previously (which we showed to be a single factor of interest design). We now outline the stages of conducting a LHC design and the subsequent computation of PRCC's.

Suppose that we have information such that we have a reasonably accurate idea of the sampling distributions of each of the k parameters in our HIV transmission model. We now split the range of each of these k distributions into n non-overlapping intervals where the probability that the parameter lies in any given one of these n intervals is $(1/n)$ for each given interval. Hence we now have n equally likely sampling intervals for each of our k parameters. For each of the k parameters we now assign an index integer from 1 to n to each of the n intervals. We now randomly select one of the equally likely intervals (noting its index number) from each of the k parameters in turn. We repeat this procedure n times without replacement. This allows us to create a matrix which contains k columns and n rows where each row represents an index number from each of the k model parameters and each column contains the indices 1 to n inclusive. For example suppose that we have three model parameters and $n = 5$. This could give us the following array of indices:

1	2	3
2	3	4
3	4	5
4	5	1
5	1	2.

The first row in this array says that we chose the interval with index number 1 for the first model parameter together with the interval with index number 2 for the second model parameter and the interval with index number 3 in the third model parameter. Once the index array has been chosen we progress through each row in turn and randomly select a parameter value from each of the equally likely intervals according to the index number chosen. In our example we would randomly select a value for the first parameter from the interval with index number 1 and similarly for the second and third model parameters. Having now chosen a random vector which contains a value for each of the k model parameters we can now compute the output variable, the long term prevalence of HIV. We repeat this procedure for each of the n rows in the index array and thus we create a new array which has $k+1$ columns (one column for each model parameter and one for the output variable) and n rows. Each row in this new array contains a value for each model parameter and the corresponding output variable. For example we may have the following array of three parameters and output variable (the final column) from the array of indices:

228.53	137.36	4.39	0.510
204.67	89.23	1.97	0.324
289.19	109.20	2.74	0.521
278.28	104.97	3.35	0.496
255.41	126.57	0.56	0.522.

It is at this stage that we now investigate the relationship between the parameter values and the output variable. Using PRCC's we examine the relationship between the columns containing the parameter estimates and the output variable.

Before we discuss the criteria which Blower and Dowlatabadi use to establish which model parameters are the most influential the design outlined above requires some comment. The Latin Hypercube design used to select the array of indices for each model parameter assumes fundamentally that the value of the output variable depends simply on a independent contribution from each model parameter. To see this it is

sufficient to note that we simply selected one treatment level from each factor in turn and computed that corresponding value of the output variable. Unlike a factorial design we do not systematically visit all permutations of the different levels of each model parameter. The Latin Hypercube design uses very little data since each parameter is sampled at each level (index number) only once. However this means that this design does not fit into the standard g.l.m. framework as it is not possible to estimate and test the significance of the individual parameters in such a model since we only have one observation at each level of each parameter (to see this note that in each column of the indices array each integer from 1 to n appears only once). For example unlike in a Latin Square Design we cannot separate the effect due to treatment level i of parameter one from treatment level j of parameter two, since these treatment levels appear only once in each parameter.

The motivation behind the Latin Hypercube design is that it can deal with a large number of parameters (factors) using the bare minimum of observations. Blower and Dowlatabadi claim that the LHC is an extremely efficient sampling scheme, however in some ways this misses the point in that it is only efficient if one assumes that the model parameters all act independently. If this is not the case then it is less clear that this design is obviously better than other more traditional experimental designs. A useful aspect of the LHC design is that it is very easy to repeat the design using an increased level of aggregation in the input parameter distributions. We shall discuss shortly why this feature is useful in order that a sensitivity analysis based on LHC produces sensible results when interaction effects exist between the model parameters. Before discussing this we now briefly outline the use of PRCC's to determine which model parameters are the most influential.

Once the data have been collected (in other words we have sampled the model parameters and calculated the corresponding values of the output variable) the next stage is to decide which parameters are the most influential. To do this we replace each of the parameter values and output variable estimates with their relative ranks amongst the other parameter values and output variable estimates. For example replacing the array given above with ranks it becomes

2	5	5	3
1	1	2	1
5	3	3	4
4	2	4	2
3	4	1	5.

Blower and Dowlatabadi examine the relationship between each model parameter and the output variable using partial rank correlation coefficients (PRCC's). In our example the PRCC for parameter one is a measure of the relationship between column one (parameter one) and column four (output variable) with any correlation between column one and columns two and three removed. The PRCC's for parameter two and parameter three have similar interpretations. The calculation of PRCC's enables the determination of the statistical relationships between each input parameter and each output variable while keeping all other input parameters constant at their expected value (Conover, 1980). This procedure enables the independent effects of each parameter to be determined, even when the parameters are correlated. A PRCC indicates the degree of monotonicity between a specific input variable and a particular output variable. The computation of PRCC's is complicated and full details are given in Appendix A in Blower and Dowlatabadi (1994). Once a PRCC has been computed for each model parameter we can then compute the following test statistic

$$t = PRCC\sqrt{\frac{n-2}{1-|PRCC|}},$$

where the distribution of t approximates a Student's t with $n-2$ degrees of freedom (Blower and Dowlatabadi, 1994). This test statistic determines the significance of a nonzero value of PRCC. In order to determine which model parameters are the most influential the number of treatment levels is fixed (the value of n) and then PRCC's are computed for each model parameter together with the value of the test statistic. Those parameters with the most extreme values of t, and hence the most statistically significant are deemed to be the parameters which are most influential in determining the output variable.

Given that the LHC design assumes that each parameter acts independently then the PRCC is a convenient way to measure the influence of each model parameter. However the problem still remains that the LHC design (and the subsequent PRCC) does not take into account interaction effects between the model parameters. We now discuss what problems this may cause to the conclusions of a sensitivity analysis.

9.5 The Value of n

We have briefly outlined the Latin Hypercube (LHC) method proposed by Blower and Dowlatabadi and we have demonstrated that this approach is analogous to a single factor of interest experimental design. We mentioned previously that a sensitivity analysis should be treated as a multiple factor experiment so is it natural to ask what problems arise from using the Latin Hypercube design in place of, for example, a full factorial design. Theoretically it seems more appropriate to use a factorial design and then use PRCC's. If an interaction effect is not present between the model parameters then the method proposed by Blower and Dowlatabadi seems very efficient, however if interaction is present this this may cause a problem. For example it may be the case that repeating the LHC with a different random array of indices may produce markedly different results, since for some combinations of parameters a large interaction effect may occur (later we illustrate this using the HIV Test Model as an example). The result of this means that it would be unwise to judge the merits of any parameter using only a single value of PRCC and its associated test statistic, as this test statistic may be unreliable since it does not incorporate the extra variability due to interaction effects. It seems intuitive that as we increase the number of treatment levels (relative to the number of parameters) in the LHC then this problem should decrease. With more observations the PRCC should be able to give a balanced account of the influence of each model parameter by taking into account parts of the sample space where interaction occurs. Hence it seems reasonable that while theoretically the LHC design is not ideal as a basis for a sensitivity analysis of a complex model of HIV transmission this can be compensated for by increasing the size of n relative to k. Blower and Dowlatabadi state that an exact formula to calculate n does not exist in the literature and the size of n for any specific analysis should be determined by the significance level required in the PRCC significance test. In light of our previous discussion in seems that it is not only the significance level which should guide the choice of n but also the amount of interaction present in any particular analysis. However this presents difficulties as it is often hard to assess the amount of interaction present.

We have illustrated that the method proposed by Blower and Dowlatabadi requires some care in choosing the size of n in order that the results are accurate. For example an indication that n is too small is if the method is repeated and PRCC's which were significant are no longer significant. We now illustrate the use of the sensitivity analysis

proposed by Blower and Dowlatabadi using the HIV Test Model.

9.6 HIV Test Model Example

We now conduct a limited sensitivity analysis on the HIV Test Model discussed in the previous chapter. We examine this model as there are good reasons for arguing that interaction effects may exist between several parameters in this model. For simplicity we assume that the only model parameters which are subject to variation are λ_1, λ_2 and δ_t. We are particularly interested in these parameters as we have already demonstrated that the effect on the long term prevalence of disease due to changing the value of λ_1 or λ_2 is heavily dependent on the current value of δ_t. Hence it would seem reasonable to expect that this model possesses a significant interaction effect between λ_1 and δ_t, and λ_2 and δ_t. Before we can conduct a sensitivity analysis we first need to estimate a sampling distribution for each of these three parameters. It was decided to use uniform distributions since we do not have any data available to estimate these distributions and for the purpose of this example the parameter distributions are of secondary importance. We decided to use $\lambda_1 \sim U(150, 300)$, $\lambda_2 \sim U(50, 150)$ and $\delta_t \sim U(0.1, 5)$, where $U(a, b)$ denotes the uniform distribution on [a,b]. Strictly speaking it may be more appropriate to use a joint distribution for λ_1 and λ_2, however for illustration purposes these simple distributions are sufficient.

We now conduct a sensitivity analysis on these three parameters. We are particularly interested in whether repeating the analysis for a different random set of parameter estimates gives us the same conclusion as to the relative importance of the parameters. Table 9.1 contains a summary of results for the HIV Test Model using only the three parameters, λ_1, λ_2 and δ_t with $n = 6$ and we have repeated the analysis for two different sets of parameter estimates (two different arrays of indices). Table 9.1 suggests that when $n = 6$ there is an interaction effect which causes problems in deciding which parameters are statistically significant from zero, in this case we have different conclusions as to the importance of λ_2 or δ_t. If we carried out only the first trial then we would conclude that λ_2 was the most influential parameter whereas the second trial suggests that λ_1 is the most important. It is important to note that while $n = 6$ is small it is not the absolute value of n which is important but the size relative to k, in this example $k = 3$. If we have more model parameters then there may exist a number of complex interaction relationships and so n must be sufficiently large to ensure that the

Table 9.1: Sensitivity Analysis with $n = 6$

Trial Number	Parameter	2.5%	t-stat	97.5%	PRCC	Significant
One	λ_1	-2.78	4.19	2.78	0.84	yes
	λ_2	-2.78	6.03	2.78	0.91	yes
	δ_t	-2.78	-0.27	2.78	-0.124	no
Two	λ_1	-2.78	13.90	2.78	0.98	yes
	λ_2	-2.78	-0.17	2.78	-0.09	no
	δ_t	-2.78	3.71	2.78	0.81	yes

sensitivity analysis takes into account these relationships when evaluating the influence of each individual parameter. It is worth noting that Blower and Dowlatabadi suggest that the LHC method is particularly useful in large, complex models, however it seems reasonable that it is these kinds of models which will contain interaction effects. Moreover as the differential equation models commonly used are non-linear it seems plausible that interaction effects will occur.

We have shown a trivial example of where the LHC method gives unreliable results. We suggested earlier that two ways of overcoming this problem are by making n sufficiently large and conducting several trials so that we can get an idea as to typical values of the test statistic. Figures 9.1-9.3 show values of the t test statistic for λ_1, λ_2 and δ_t for $n = 30$ with the sensitivity analysis repeated 8 times. The horizontal lines on the plot are the upper and lower significance limits at 97.5% and 2.5% respectively for a Student's t distribution with 28 degrees of freedom. From these figures it is clear that the PRCC's for λ_1 and λ_2 are statistically different from zero when $n = 30$, it is less clear from Figure 9.3 whether this is the case for δ_t. While obviously the value of the test statistic is subject to variation the spread of values of this statistic gives a good idea as to which model parameters are the most influential. Examining the values of the test statistics in the figures it seems reasonable to conclude that out of the these three parameters λ_2 is the most influential closely followed by λ_1 and finally δ_t. We mentioned earlier that rather than repeat the sensitivity analysis a number of times we could simply increase n and use a single trial. The problem with this method is that we do not know how large to choose n such that we draw the correct conclusion. For example with $n = 100$ in the HIV Test Model we still have that one out of ten test

Figure 9.1: Test Statistic for λ_1

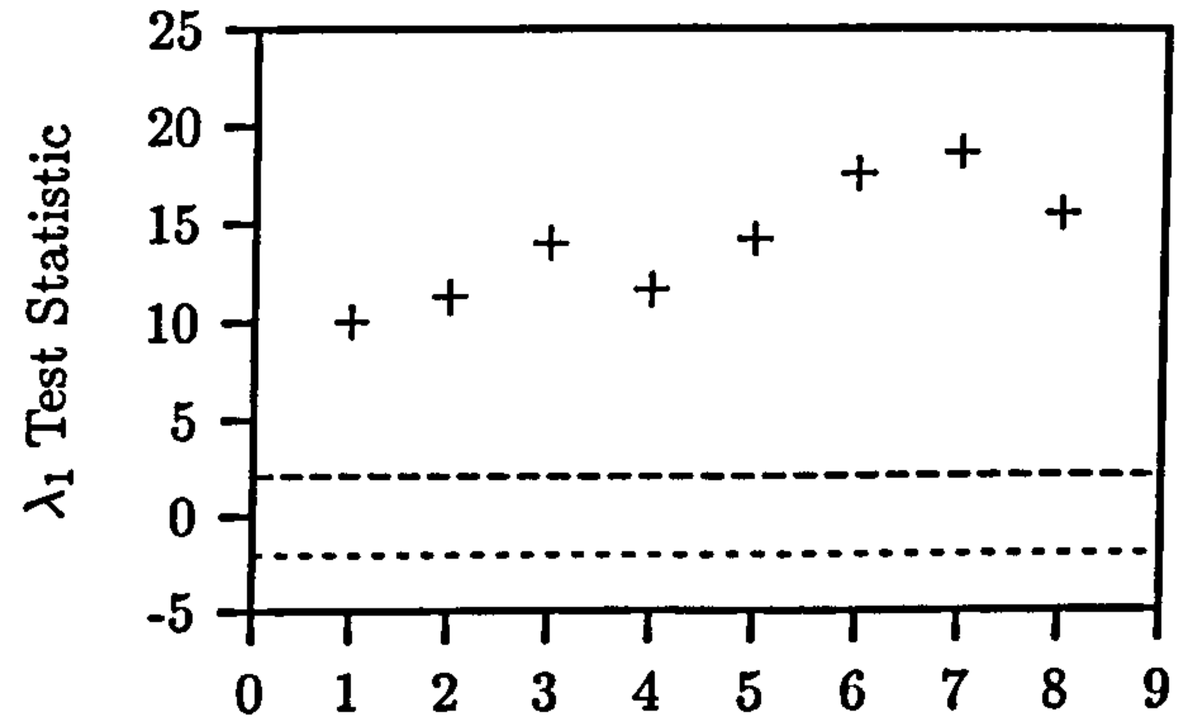


Figure 9.2: Test Statistic for λ_2

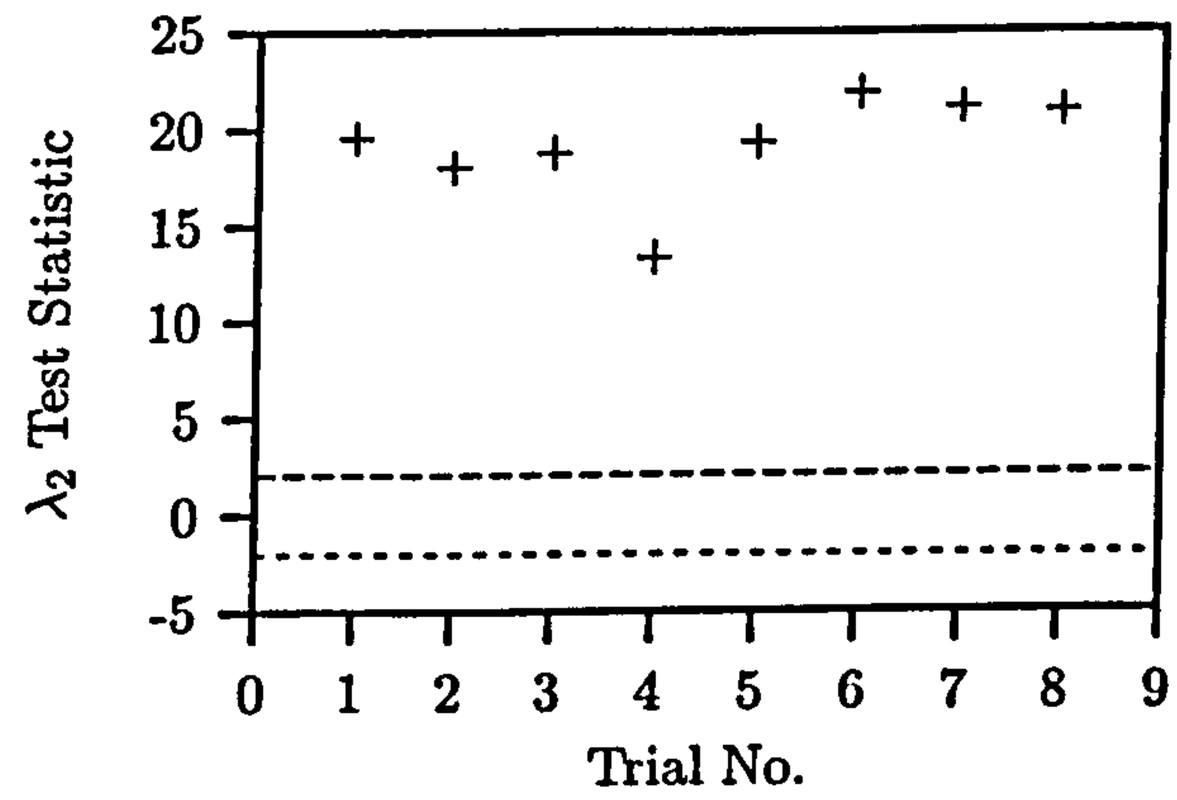
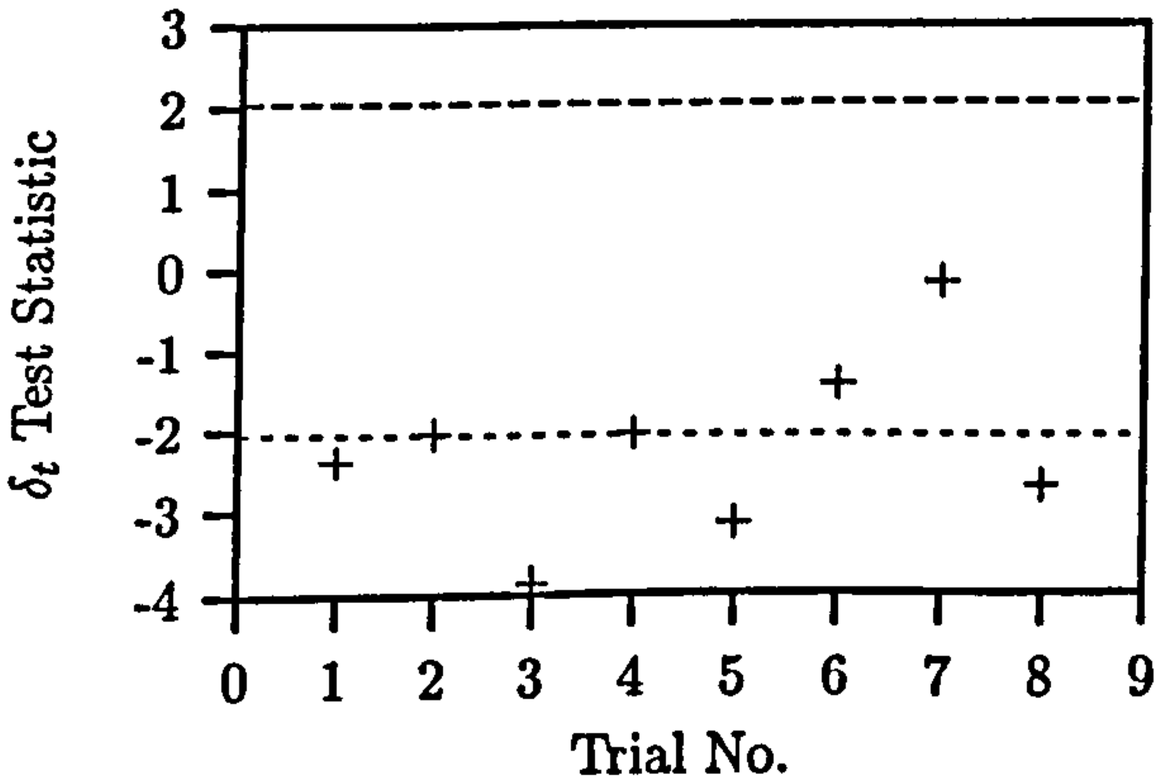


Figure 9.3: Test Statistic for δ_t



statistics still gives the wrong conclusion (at a 5% significance level) as to the relative importance of λ_1 and λ_2. From the previous figures it seems reasonable that we should conclude that λ_2 is the most influential parameter. Hence it may be the case that it is better at any value of n to repeat the analysis a number of times to avoid misleading conclusions.

9.7 Summary of Sensitivity Analysis

In this chapter we have examined in detail the method of sensitivity analysis proposed by Blower and Dowlatabadi (1994). We first discussed the similarities between experimental trials and a sensitivity analysis, and outlined several common experimental designs and commented on their relation to the Latin Hypercube (LHC) design proposed by Blower and Dowlatabadi. We stated that on theoretical grounds the LHC design was not ideal for a sensitivity analysis as this design assumes that the affect of changing one model parameter is independent of the values of the other model parameter. In other words the LHC design does not take into account the interaction effects between the different model parameters. We used the HIV Test Model from Chapter 8 to illustrate the potential problems of using the LHC design together with partial rank correlation coefficients (PRCC's) when an interaction effect is present, and we showed that this can lead to unreliable conclusions as to which are the most influential model parameters. We next discussed ways of using the LHC design such that sensible results are obtained when interaction effects are present. We suggested that a suitable method was to repeat the LHC design a number of times, and compare the resulting PRCC t-test statistics for the various model parameters. In general the parameters which have the most extreme test statistics are those which are most influential in affecting the output variable.

Chapter 10

Stochastic Models of Needle Sharing

All the models we have examined in Chapters 2-8 have been deterministic in nature. A main advantage of using deterministic models (as opposed to stochastic models) is that they are generally more tractable than the stochastic alternative. The flip-side of this is that a deterministic approach may not capture the true behaviour of an epidemic. The real world is highly stochastic, particularly where the behaviour of humans is concerned. In this penultimate chapter we examine a number of comparisons between the models we have studied previously and their more realistic stochastic equivalents. We are interested in whether the conditions necessary for the disease to take off or die out are the same as in the deterministic case and in particular whether the long term prevalence of disease is comparable.

We first discuss several fundamental differences between stochastic and deterministic models. We then examine a comparison of the Kaplan and O'Keefe Model defined by eqns (1.1)-(1.2) with its stochastic equivalent. This leads on to a study of the probability of extinction in stochastic needle sharing models. We next briefly compare the Simple, Optimistic and Pessimistic Models with their stochastic equivalents. Finally we examine a stochastic alternative to the HIV Test Model and its extension to three stage infectivity. The chapter concludes with a short summary.

10.1 Stochastic and Deterministic Models

There are two fundamental questions regarding the spread of an infectious disease. Firstly, what is the likelihood that an epidemic will occur, and secondly, if an epidemic does occur what is the likely impact on the population? Epidemiological models can be valuable in providing answers to these questions. Deterministic models are used extensively in epidemiology and have been shown to provide useful insight in many areas. Their use is commonly justified by arguing that this approach is a good approximation to the equivalent stochastic process when the population size is large. However except in trivial cases there are no analytical results to support the similarity of deterministic and stochastic models.

Stochastic and deterministic models provide different answers to the two questions posed above. Suppose for example that we are in the common situation where a critical threshold exists between the disease dying out or taking off. In a deterministic model, if the estimates of the model parameters give rise to a value greater than the critical threshold, then if initially present in at least one addict or one needle the disease will take off. However in a stochastic model (which allows for the random nature of disease transmission) even if the parameters give a value greater than the critical threshold an epidemic may not occur and the disease could simply infect a small number of individuals before dying out. This is an important distinction between deterministic and stochastic models.

We now turn to the case where an epidemic does occur and we wish to estimate the long term prevalence of disease. It is possible for deterministic models to reach a state of equilibrium where disease remains forever present in the population. This may be through the disease remaining at a constant level equal to an endemic equilibrium solution, (we have demonstrated that this situation is approached for all the models we have so far discussed). Alternatively it may be that the disease is always present and exhibits cyclical behaviour. In either case the deterministic system can be thought of as having reached equilibrium. This is not so in a stochastic system, where the system usually represents a continuous time Markov chain. Obviously in such a situation we have that the disease-free state is an absorbing state, hence this zero state is persistent non-null (the probability of the chain eventually reaching this state equals unity and the expected time for this to occur is finite), and all other states are transient (the probability of the chain ever returning to one of these states given that it starts in that

state is strictly less than unity). This chain is not irreducible but there exists a unique stationary distribution where no addicts and no needles are infected.

The Markov chain described above can be thought of as a fluctuating sequence denoting the number of people infected at time t. Either this sequence becomes so large that it escapes to infinity or it is absorbed at zero during one of its fluctuations. Since we are interested in the application of this chain to the spread of disease among a real population, we have only a finite state space and therefore the number of people infected by the disease cannot become so large that it escapes to infinity. Therefore we have that eventually the disease must die out. For example if we were to simulate such a chain 100 times then after a sufficiently large duration we expect that each of these simulations will have been absorbed at the disease-free state.

At the start of an epidemic when the number of infected addicts and infected needles is small then we expect the probability that the chain will reach the absorbing state to be particularly high. However once the disease has become more established (for example after a stage of exponential growth) then it may settle down to a quasi-equilibrium state (Renshaw, 1991, Hay, 1999), this is where the disease appears to fluctuate about a steady-state. For the reasons mentioned above this is not a true equilibrium state (or stationary distribution), however as discussed by Hay (1999) and Renshaw (1991) the time taken for the chain to reach the absorbing state once this quasi-equilibrium state has been reached may be biologically irrelevant to the system under consideration (for example it may take thousands of years). This means that comparing the quasi-equilibrium of a Markov chain with an equilibrium solution of an equivalent deterministic model is a sensible comparison.

As a remark on the merits of using stochastic models over deterministic models or vice versa, Mollison (1991) suggests the use of linearised deterministic models and non-linear stochastic models. In this paper Mollison specifically examines population velocities in various types of spatial epidemic models, however a number of the points he raises carry over to non spatial models. Essentially he argues that the spread of disease in a non-linear differential equation model should be bounded above by its linearisation. Therefore simple linearised models provide a convenient basis from which to investigate the criteria necessary for an epidemic to occur. Mollison argues that (at least in a spatial context) non-linear stochastic models should be used to investigate the long term behaviour of disease, as non-linear deterministic models offer little advantage over linear deterministic and linear stochastic models.

As mentioned above non-linear stochastic and non-linear deterministic models need not behave in a similar fashion, for example they may not suggest the same long term prevalence of disease. Therefore it is of interest to establish whether the deterministic models which we have spent considerable time examining in the previous chapters behave in a similar fashion to their more realistic stochastic equivalents.

10.2 Simulation Aspects

Stochastic simulation is a common technique, and as such we do not give in depth details of the practical workings of the simulation method. However a brief discussion is pertinent. The computer based simulations in this chapter were written using a computer program in C. In essence this program treats each member of the population (each addict) as a separate entity who makes his or her own decisions independently of any other addict. Each addict is allowed to make a number of decisions, for example when he or she will next inject with a shared needle and whether or not they will clean this needle prior to use. In addition certain external forces act on each of the addicts, for example addicts can be removed from the population and can develop full blown AIDS (at which point they also leave the population). We feel that this method of simulating the population is as good a mimic of real life as possible within the framework of our model assumptions.

In terms of practicalities the simulation programs in this section are very computationally intensive and require considerable computing time. For this reason all simulations illustrated in this chapter use a relatively small population size of 91 addicts and 100 needles (this ensures we have the "gallery ratio" as estimated in Table 2.1). The computation time for 100 simulations of this population size for a duration of 200 years is between 12 and 36 hours (for most of the models in this chapter). Using larger populations increases this time considerably, and we expect that for large populations the stochastic model is close to the behaviour of the corresponding deterministic model, hence there is little advantage in increasing the addict population size. However most models in this chapter were simulated over 400 years using population sizes from 50 to 300 addicts in order to identify whether the behaviour of the models was dependent on population size. Simulations suggest that this is not the case and a population size of 91 addicts and 100 needles is representative of the behaviour of the disease among larger populations. It should be noted that since we are using a population size of 91

addicts the simulations in this chapter are obviously not as smooth as the numerical approximations to the deterministic models of previous chapters. The smallest change in prevalence in the stochastic models is 1/91 compared to a precision of 16 significant digits in the deterministic models.

10.3 The Kaplan and O'Keefe Model

Bearing in mind the discussion in Section 10.1 we now compare the behaviour of the Kaplan and O'Keefe Model with its stochastic equivalent. We do not aim to present a rigorous comparison between stochastic and deterministic models. We simply focus on whether the deterministic Kaplan and O'Keefe Model provides a good approximation to the behaviour of a simulated stochastic epidemic based on the same behavioural assumptions. Using Assumptions 1 to 9 in Section 2.2 it is straightforward to construct a computer-based simulation model which represents a stochastic equivalent of the deterministic Kaplan and O'Keefe Model.

Figure 10.1 shows a single simulation of the fraction of addicts infected over time in the stochastic model for a population size of 91 addicts and 100 needles with initially only a single infectious addict in the population. The figure also shows a numerical simulation of the Kaplan and O'Keefe Model (using the same parameter estimates as in the stochastic model) where initially a fraction 0.011 (1/91) of the addict population was infectious and no needles were initially infectious. The value of R_0 for both these models is 3.19. It could be argued that after the initial growth stage the stochastic realisation does tend to fluctuate about the endemic equilibrium level of the Kaplan and O'Keefe Model. While we have illustrated the behaviour of the stochastic model over a duration of only 150 years, simulations of this model over 600 years show similar fluctuations about the endemic equilibrium level of the deterministic Kaplan and O'Keefe Model. Moreover we find that increasing the population size to three hundred addicts (with the ratio of addicts to needles kept constant) does not appear to decrease the amplitude of these stochastic fluctuations. We do not illustrate simulations of the Kaplan and O'Keefe Model when $R_0 < 1$ but we find that the disease dies out in all addicts and all needles and moreover this occurs on a biologically realistic time frame (less than 200 years). This is good evidence that as in the deterministic model when $R_0 < 1$ the disease dies out among the population.

Figure 10.2 was constructed by simulating the stochastic model 100 times using the

Figure 10.1: Single Simulation of Stochastic Model ($R_0 > 1$)

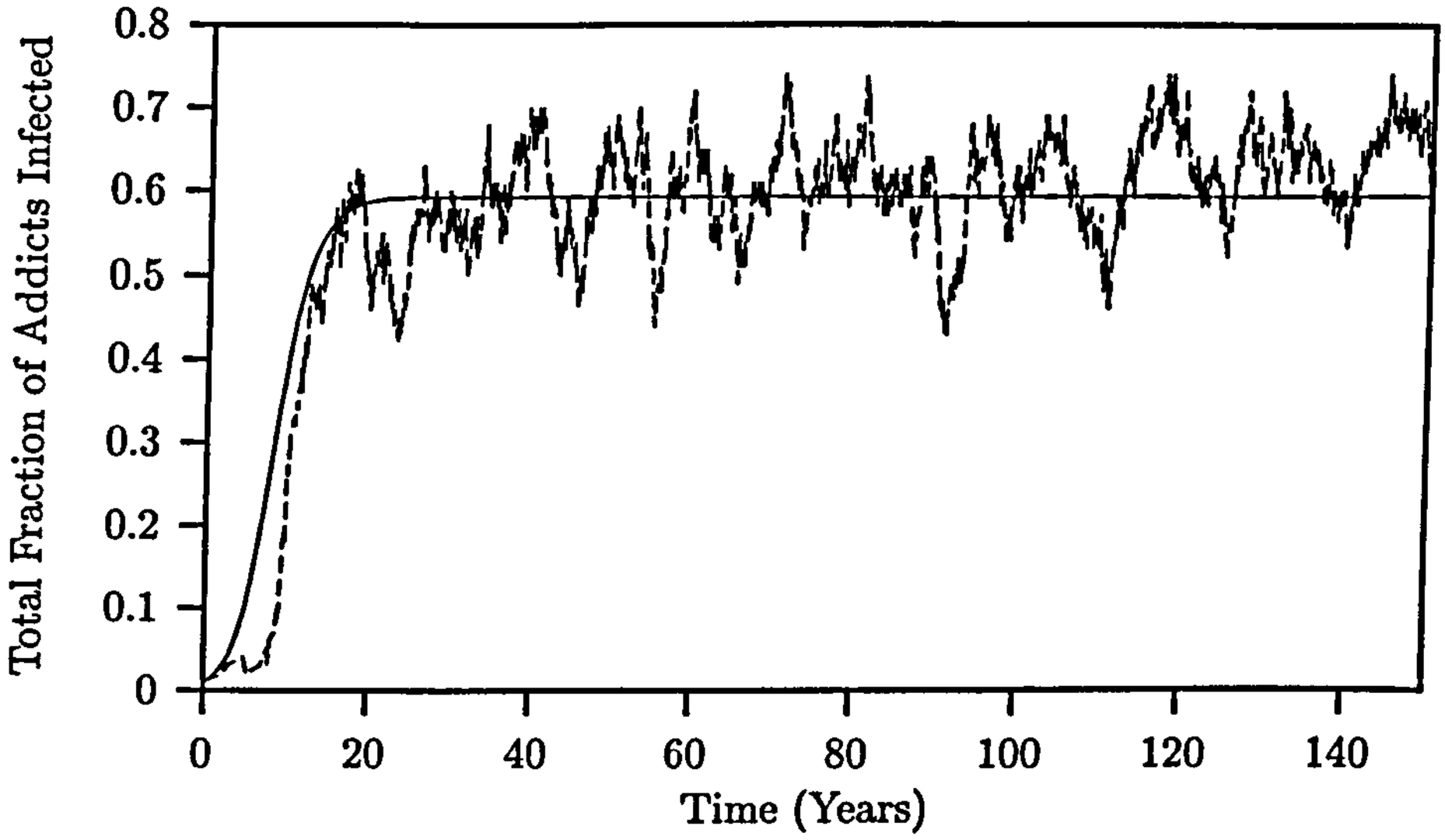
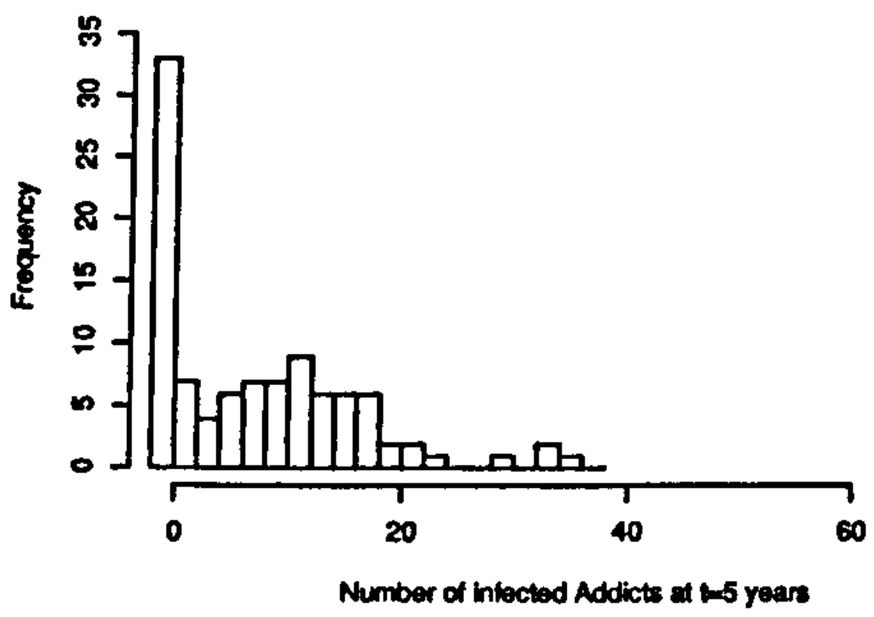
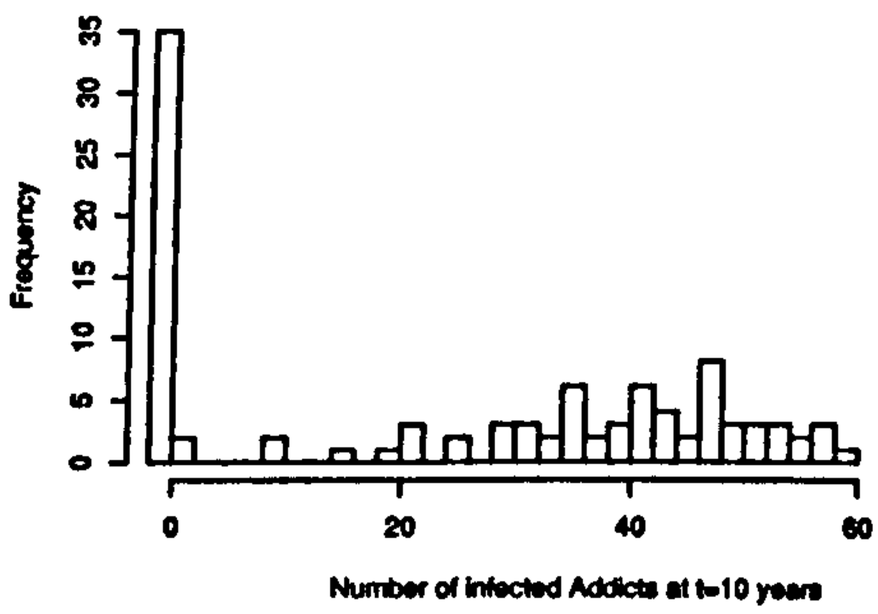
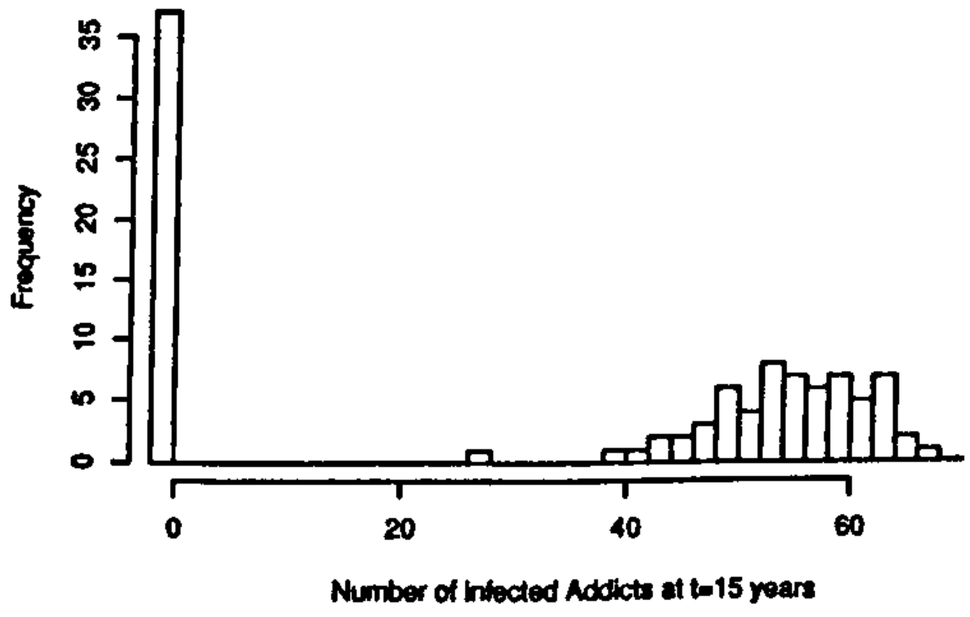
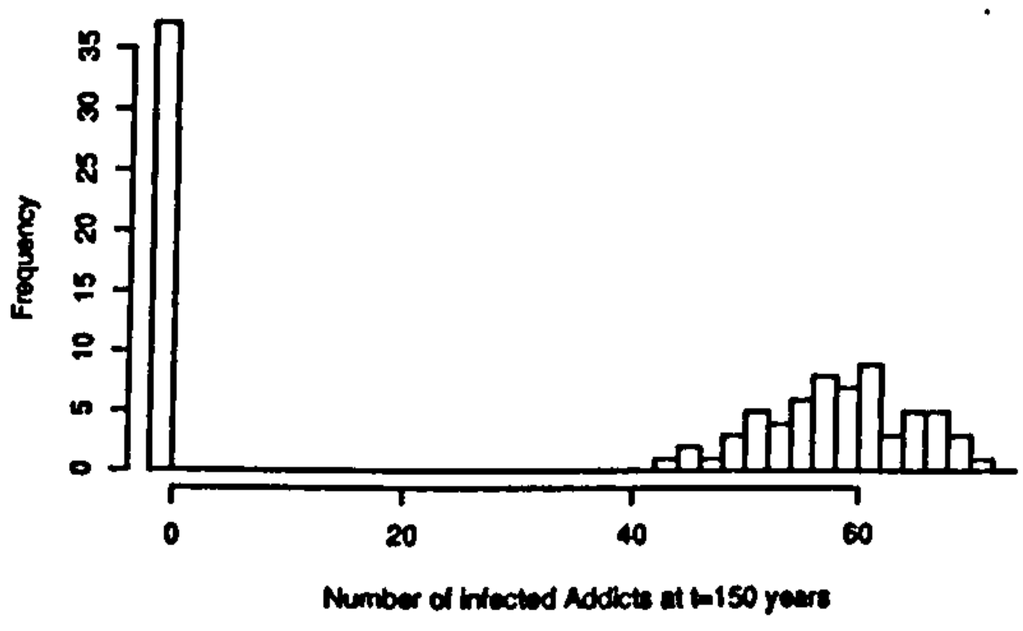


Figure 10.2: 100 Simulations of Stochastic Epidemic: One Addict Initially Infectious







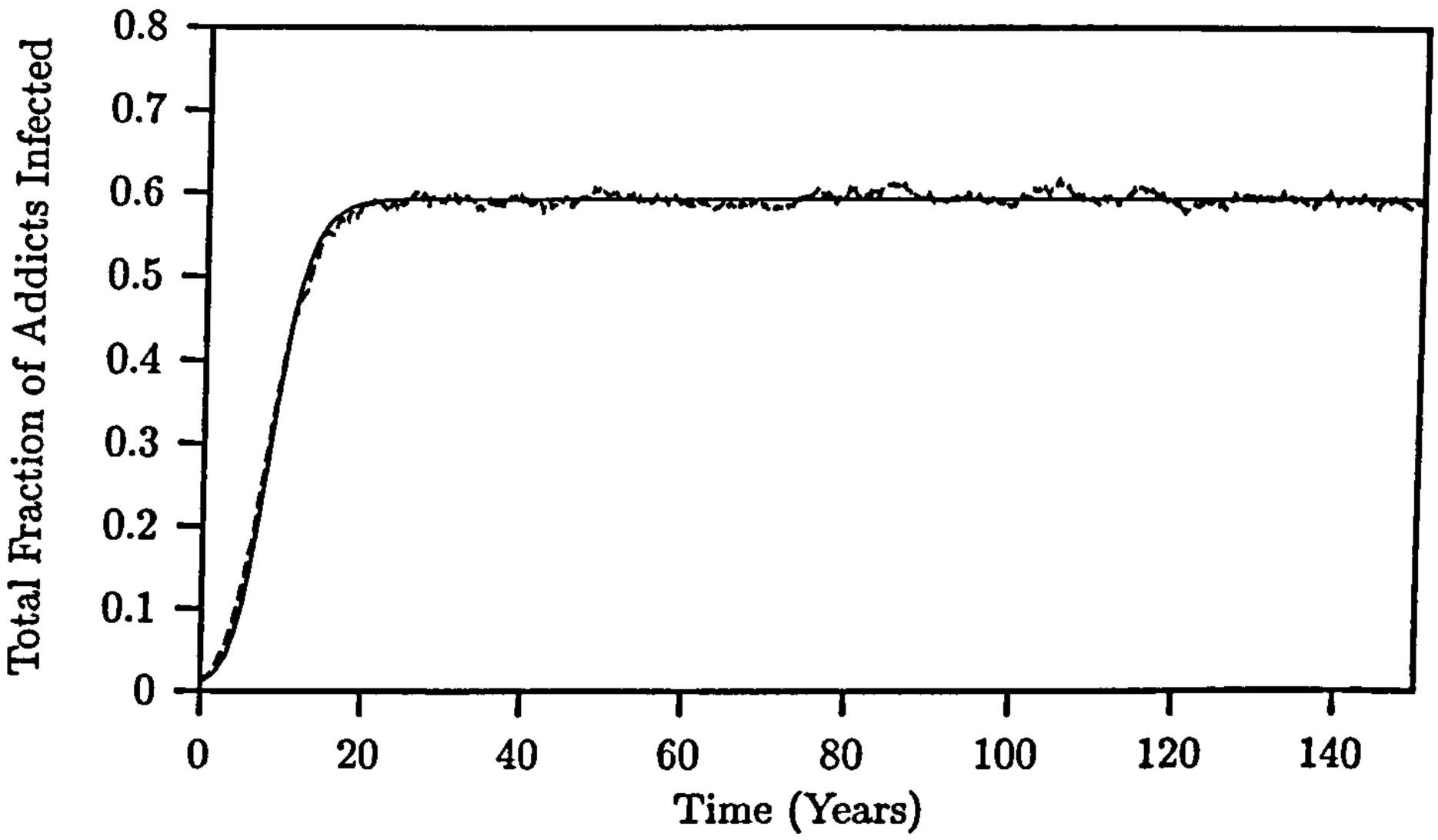


same parameter estimates as in Figure 10.1 and in each simulation we stored the number of infectious addicts at $t = 5$, 10, 15 and 150 years. It was initially assumed that only a single addict was infectious and no needles were initially infectious and the total (fixed) population size was 91 addicts and 100 needles. The figure shows that even though $R_0 = 3.19 > 1$ the disease dies out in a minority of simulations. After approximately 15 years has passed no more simulations die out, this suggests that if disease is still present after 15 years then it will tend an endemic state. However given our previous comments this is not a true equilibrium state but rather a quasi-equilibrium state.

It appears that the criteria for the disease to take off is the same in the Kaplan and O'Keefe Model and its stochastic equivalent, however as previously mentioned the disease can also die out in the stochastic case. We are now interested in whether the long term prevalence of disease given that an epidemic does actually occur is similar to that of the deterministic Kaplan and O'Keefe Model. An appropriate method of comparing these models is to compare the expected value of the stochastic model conditional on the disease not dying out, with the long term prevalence in the Kaplan and O'Keefe Model. It seems more appropriate to use conditional expectation since the long term prevalence of disease is of interest only provided that an epidemic does in fact occur. Indeed the motivation for studying HIV and AIDS models is that this disease has been shown to take off among particular population classes, for example among intravenous drug users.

Figure 10.3 shows the average fraction of addicts infected out of $N = 100$ simulations of a population of 91 addicts and 100 needles conditional that the epidemic takes off. The initial condition in each simulation was one infectious addict and no infectious needles. This expected value is based on 63 out of 100 simulations since in 37 of these the disease dies out (as shown in Figure 10.2). From the figure it appears that the Kaplan and O'Keefe Model is a very good approximation of the conditional mean of the stochastic process and not only in terms of long term prevalence, the dynamic behaviour is also very similar. For larger population sizes (not illustrated) the Kaplan and O'Keefe Model is also an excellent approximation to the conditional mean of the stochastic equivalent.

Figure 10.3: Kaplan and O'Keefe Model Comparison



10.3.1 Probability of Extinction in the Kaplan and O'Keefe Model

In the previous section we demonstrated that the disease dies out in a minority of simulations when $R_0 > 1$. We now use a similar argument to that used by Bartlett (1955) to derive an expression for the probability of extinction of disease (or equivalently the probability of a major outbreak) based on the modelling assumptions made by Kaplan and O'Keefe (1993) which we discussed in detail in Chapter 2.

Consider the general stochastic epidemic model without vital dynamics, where initially there are n susceptibles and a infectives, and a is small. Bartlett (1955) points out that when n is large, the population of infectives is approximately subject to a birth and death process with birth and death rates βn and γ respectively. Now the chance of ultimate extinction for such a process is R_0^{-a} if $R_0 > 1$ and 1 if $R_0 \leq 1$, where $R_0 = \beta n/\gamma$, γ is the per capita removal rate and β is the rate at which each infected individual and each susceptible individual make potentially infectious contacts so that if there are x susceptibles and y infectives then the transmission term is βxy. Hence intuitively we expect a minor epidemic outbreak to occur with probability R_0^{-a} and a major epidemic outbreak with probability $1 - R_0^{-a}$ if $R_0 > 1$. For $R_0 \leq 1$ the epidemic outbreak is always minor.

Whittle (1955) makes this intuitive argument more rigorous by bounding the process above and below by two birth and death processes for which the probability of extinction

can be explicitly calculated. The ideas of Whittle's argument are explained clearly by Bailey (1975). If π_i denotes the chance that not more than a proportion i of the n susceptibles are eventually attacked then Whittle shows that:

(i) If $R_0 > (1-i)^{-1}$, then $R_0{}^{-a} \leq \pi_i \leq (R_0/(1-i))^{-a}$;

(ii) If $(1-i)^{-1} > R_0 > 1$, then $R_0{}^{-a} \leq \pi_i \leq 1$;

(iii) If $R_0 \leq 1$, then $\pi_i = 1$.

In this section we outline a similar intuitive branching process argument to calculate the probability of a major epidemic outbreak for our model of HIV among injecting intravenous drug users. Our method is based on a similar approximation to that suggested by Bartlett in that the start of our epidemic is approximated by a branching process which ignores small initial changes in the number of infected addicts and the number of infected needles. We do not obtain a rigorous bound for this probability as in Whittle's Theorem. We feel that it would be possible to do this using Whittle's method but the details would be very technical and we wish to concentrate on the practical result. Recall that $R_0 = \lambda(1-\phi)\alpha/((\mu+\delta)(\hat{\theta}+\hat{\tau}))$. We define $p = \alpha(1-\hat{\theta})/(\hat{\tau}+\hat{\theta}+\alpha(1-\hat{\theta}))$. If we consider a single infectious needle entering a population at the disease-free equilibrium then p represents the probability that at least one addict who does not flush the needle is infected by it.

Theorem 10.1 *Suppose that there are initially $n+a$ addicts of whom a are infected and $m+b$ needles of which b are infected and n and m are large and a and b are small. Then a large outbreak of HIV will occur with probability zero if $R_0 \leq 1$, and probability $1-(P_A^a P_B^b)$, if $R_0 > 1$, where*

$$P_A = \frac{1}{p + R_0(1-p)} \qquad \text{and} \qquad P_B = 1 + \frac{(\mu+\delta)}{\lambda}(1-p)(1-R_0).$$

Proof.

Let X denote the number of needles infected by a single infectious addict during his or her infectious lifetime in an otherwise entirely susceptible addict population with no infectious needles $(a = 1, b = 0)$, Z denote the number of secondary addicts infected in the same situation and Y be the number of addicts infected by a single infectious needle during its infectious lifetime in an entirely susceptible addict population with no other infectious needles $(a = 0, b = 1)$. Denote the probability generating functions (p.g.f.'s) of X, Y and Z by $G_X(s)$, $G_Y(s)$ and $G_Z(s)$ respectively.

We have that

$$\begin{aligned} P(X=0) &= 1-q, \\ P(X=1) &= q(1-q), \end{aligned}$$

so in general $P(X=x) = q^x(1-q)$ for $x=0,1,\ldots,$ where $q=\dfrac{\lambda}{\mu+\delta+\lambda}$.

Therefore

$$G_X(s) = E(s^X) = \frac{1-q}{1-qs}.$$

We now derive an expression for $G_Y(s)$. Each time that a susceptible addict uses an infectious needle he or she is either infected by it or not, and the needle is either flushed or not. For simplicity we assume that flushing and HIV transmission are independent, this is not necessarily true but (as discussed previously) Greenhalgh and Hay (1997) demonstrated that this assumption does not affect the number of secondary infections. Since we are assuming the Markov property in all addict-needle interactions the state of a needle after use can be treated as a renewal process.

Firstly consider the probability that an infectious needle does not infect any addicts during its infectious lifetime. A needle can be exchanged or cleaned prior to use, in either case the needle will fail to infect any addicts, this occurs with probability $(\phi+\hat{\tau})/(1+\hat{\tau})$. The probability that a needle is neither cleaned nor exchanged prior to use and therefore still infectious at the start of the injection process is $(1-\phi)/(1+\hat{\tau})$. During injection transmission does not occur and the needle is flushed with probability $(1-\alpha)\theta$, or transmission does not occur and the needle is not flushed with probability $(1-\alpha)(1-\theta)$. Therefore denoting $P(Y=y)=P_y$ we have that

$$P_0 = \frac{(1-\phi)}{1+\hat{\tau}}\Big[(1-\alpha)\theta+(1-\alpha)(1-\theta)P_0\Big]+\frac{\phi+\hat{\tau}}{1+\hat{\tau}}.$$

Solving this equation for P_0 we have that

$$P_0 = 1-\frac{c}{1-p},$$

where

$$p = \frac{\alpha(1-\hat{\theta})}{\hat{\tau}+\hat{\theta}+\alpha(1-\hat{\theta})}$$

and

$$c = \frac{\alpha(1-\phi)(\hat{\tau}+\hat{\theta})}{(\hat{\tau}+\hat{\theta}+\alpha(1-\hat{\theta}))^2}.$$

Moving on to P_1 we require that a total of one addict is infected by a single infectious needle during its infectious lifetime. Hence we require that the needle is neither cleaned

nor exchanged prior to the current injection, this occurs with probability $(1-\phi)/(1+\hat{\tau})$. Therefore in a similar fashion to above we find that

$$P_1 = \frac{(1-\phi)}{1+\hat{\tau}}\Big[\alpha\theta + \alpha(1-\theta)P_0 + (1-\alpha)(1-\theta)P_1\Big].$$

Solving this equation for P_1 we find that

$$P_1 = c.$$

It follows similarly that the probability of exactly two addicts becoming infected by a single infectious needle during its infectious lifetime is

$$P_2 = \frac{(1-\phi)}{1+\hat{\tau}}\Big[\alpha(1-\theta)P_1 + (1-\alpha)(1-\theta)P_2\Big].$$

Solving this equation we deduce that

$$P_2 = cp.$$

In general we find that

$$P_0 = 1 - \frac{c}{1-p}$$

and $$P_y = cp^{y-1}, \qquad \text{for} \quad y = 1,2,\ldots,$$

from which it is straightforward to show that

$$G_Y(s) = E(s^Y) = 1 - \frac{c}{1-p} + \frac{cs}{1-ps}.$$

We have so far derived the p.g.f for X, the number of needles infected by a single infectious addict and Y, the number of addicts infected by a single infectious needle. Now suppose that $a = 1$ and $b = 0$ so that initially there is only a single infectious addict in the population and all other addicts and needles are uninfectious. During his or her time in the population this single infectious addict infects $X = x$ uninfectious needles for $x = 0,1,\ldots$. The probability generating function of Z, the number of secondary addicts infected from this single infectious addict is

$$\begin{aligned} G_Z(s) &= E(s^Z), \\ &= \sum_{x=0}^{\infty} E(s^Z|X=x)P(X=x), \\ &= \sum_{x=0}^{\infty} E(s^{Y_1+Y_2+\ldots+Y_x}|X=x)P(X=x), \end{aligned}$$

where Y_i is the number of secondary addicts infected by the i'th needle. As the Y_i's are independent and identically distributed:

$$\begin{aligned} G_Z(s) &= \sum_{x=0}^{\infty} \left(E(s^Y)\right)^x P(X=x), \\ &= \sum_{x=0}^{\infty} G_Y(s)^x P(X=x), \\ &= E\left[G_Y(s)^X\right], \\ &= G_X(G_Y(s)), \\ &= \frac{1-q}{1-qG_Y(s)}, \\ &= \frac{1-q}{1-q\left(1-\frac{c}{1-p}+\frac{cs}{1-ps}\right)}. \end{aligned}$$

The probability of extinction is the smallest root, $s_A^\star$, of $s = G_Z(s)$ in $[0,1)$. The equation

$$s = G_Z(s)$$

can be written as

$$s - qs\left(1-\frac{c}{1-p}+\frac{cs}{1-ps}\right) = 1-q.$$

This can be re-expressed as the quadratic

$$s^2\left(cq+p\left(1-q+\frac{cq}{1-p}\right)\right) - s\left(p(1-q)+\left(1-q+\frac{cq}{1-p}\right)\right)+1-q = 0.$$

Since $s_1^\star = 1$ must be a root we find that the other root is

$$\begin{aligned} s_2^\star &= \frac{1-q}{cq+p\left(1-q+\frac{cq}{1-p}\right)}, \\ &= \frac{(1-q)(1-p)}{cq+p(1-p)(1-q)}, \\ &= \frac{1}{p+\frac{cq}{(1-p)(1-q)}}, \\ &= \frac{1}{p+R_0(1-p)}, \end{aligned}$$

as $R_0 = E(X)E(Y) = cq/((1-p)^2(1-q))$. Hence if $R_0 \leq 1$ and there is initially only one infectious addict in the population then $s_A^\star = 1$ and the disease dies out in all addicts and all needles. If $R_0 > 1$ then under the branching process approximation the disease dies out in all addicts and all needles with probability $s_A^\star = s_2^\star < 1$.

Now consider the case where initially only a single needle is infectious and all addicts are susceptible. During the time for which this single needle is infectious it infects $Y = y$

susceptible addicts for $y = 0, 1, 2, \ldots$. Let $\tilde{Z}$ denote the number of secondary infectious needles attributable to this single infectious needle. Then

$$\begin{aligned} G_{\tilde{Z}}(s) &= E(s^{\tilde{Z}}), \\ &= \sum_{y=0}^{\infty} E(s^{\tilde{Z}}|Y=y)P(Y=y), \\ &= \sum_{y=0}^{\infty} E(s^{X_1+X_2+\ldots+X_y}|Y=y)P(Y=y), \end{aligned}$$

where X_i is the number of secondary needles infected by the i'th addict. As the X_i's are independent and identically distributed, in a similar fashion to previously we have that:

$$\begin{aligned} G_{\tilde{Z}}(s) &= G_Y(G_X(s)), \\ &= 1 - \frac{c}{1-p} + \frac{c(1-q)}{1-qs-p(1-q)}. \end{aligned} \tag{10.1}$$

The probability of eventual extinction of disease is the smallest root, $s_B^\star$, of $s = G_{\tilde{Z}}(s)$ in $[0,1)$. Solving $s = G_{\tilde{Z}}(s)$ using $G_{\tilde{Z}}(s)$ from eqn (10.1) we require the smallest root of

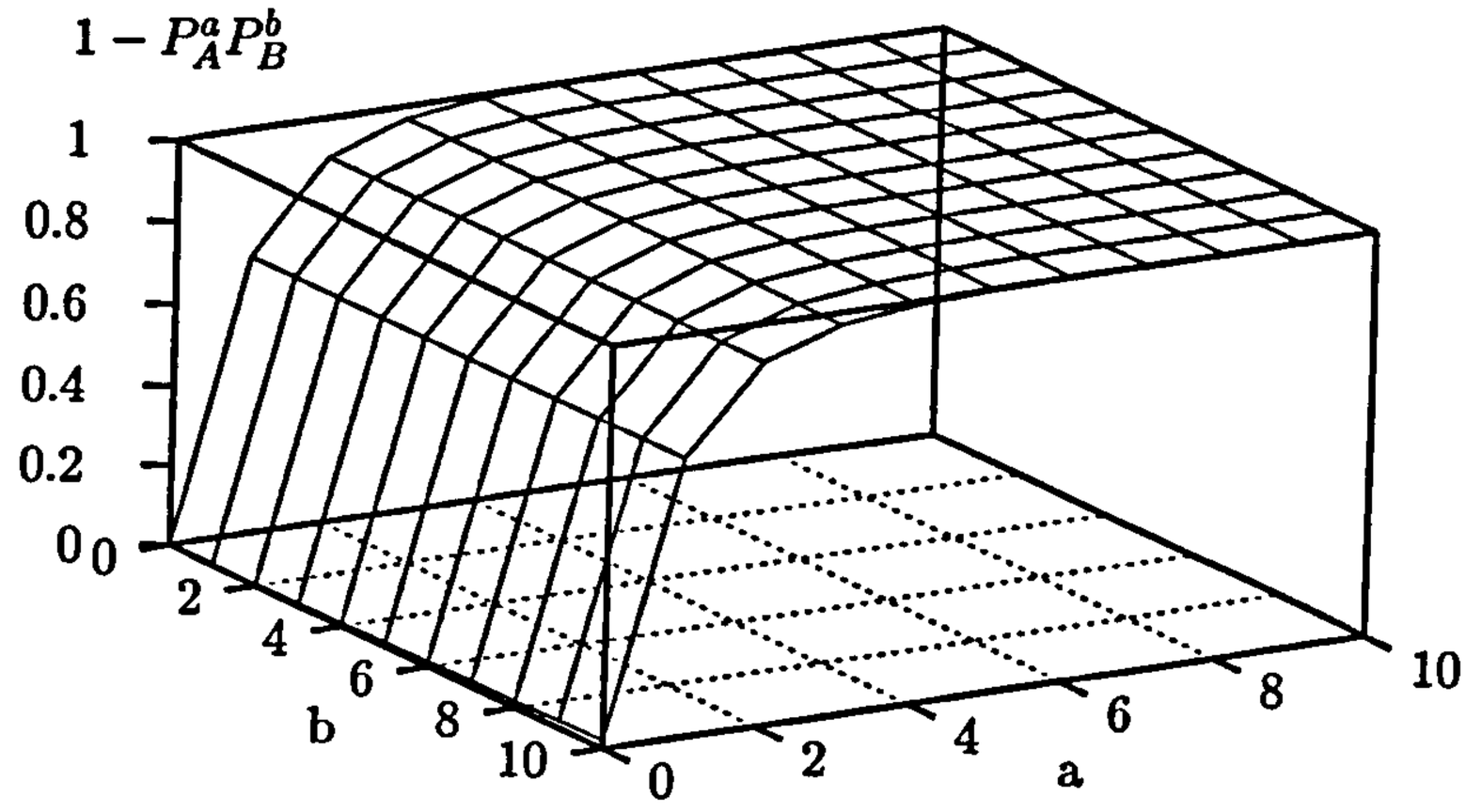
$$qs^2 + s\left(p(1-q) - 1 - q + \frac{cq}{1-p}\right) + 1 - p(1-q) - \frac{cq}{1-p} = 0.$$

Since we again have that $s_1^\star = 1$ is a root it follows directly that the other root is

$$\begin{aligned} s_2^\star &= \frac{1}{q}\left[1 - p(1-q) - \frac{cq}{1-p}\right], \\ &= 1 + \frac{(\mu+\delta)(1-p)(1-R_0)}{\lambda}. \end{aligned}$$

Hence if $R_0 \leq 1$ and there is initially only one infectious needle in the population then $s_B^\star = 1$ and the disease dies out in all addicts and all needles. If $R_0 > 1$ then the disease dies out in all addicts and all needles with probability $s_B^\star = s_2^\star < 1$. If initially there are n susceptible and a infected addicts and m infected and b uninfected needles, where a and b are small and n and m are large then under the branching process approximation we can regard the a infected addicts and b infected needles as starting independent branching processes and the probability that the disease dies out in all of them is $(s_A^\star)^a(s_B^\star)^b$. This completes the proof of Theorem 10.1.•

Figure 10.4: $1 - P_A^a P_B^b$ for varying a and b (Kaplan and O'Keefe Model)



Practical Implications

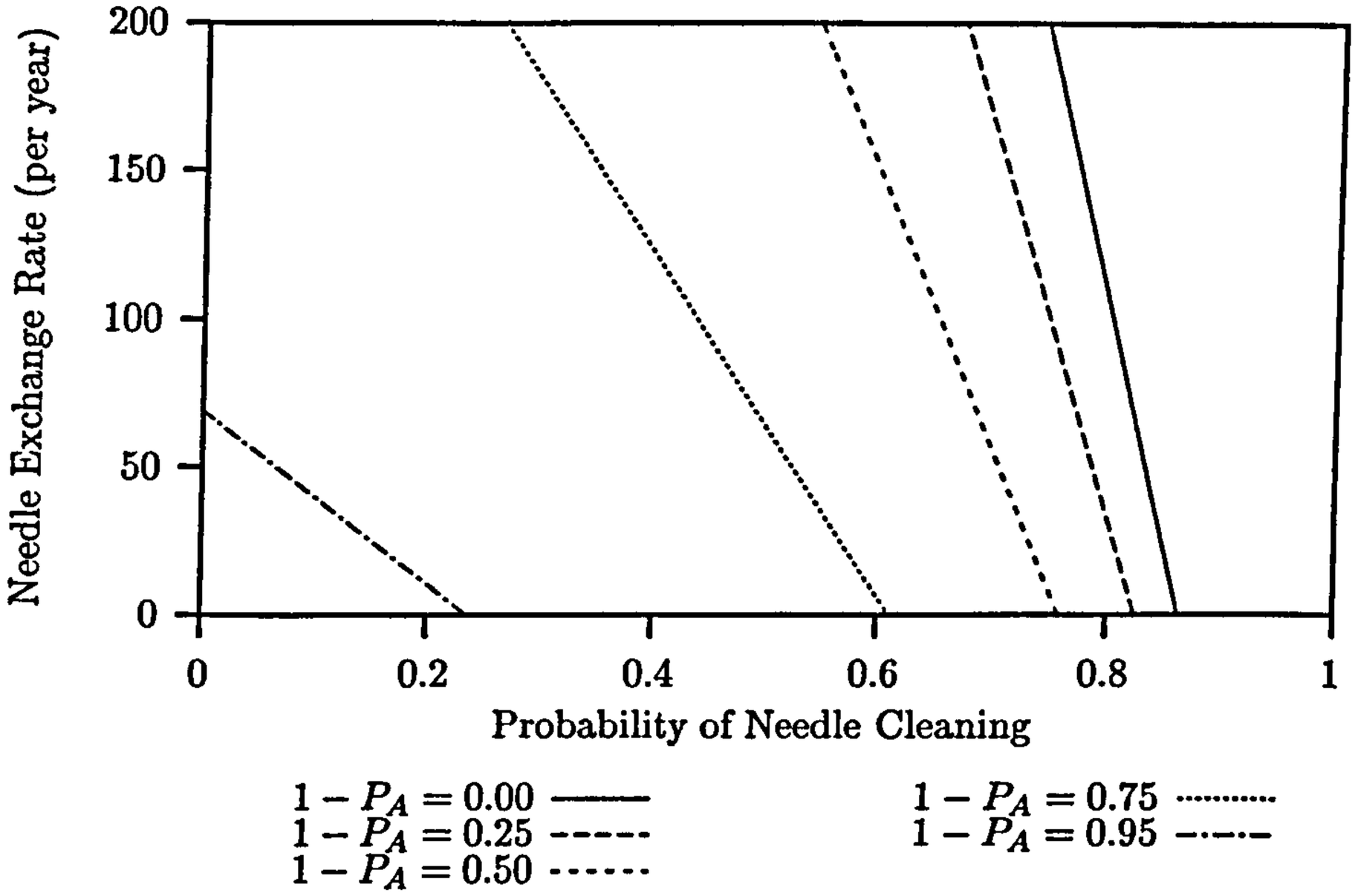
Having derived expressions for the probability that a large outbreak of HIV will occur we now examine which factors are the most significant in contributing towards such an outbreak. We first illustrate the importance of the initial conditions in our model, in particular the difference between when addicts are initially infectious and when needles are initially infectious. In the subsequent figures we use the parameter estimates detailed previously in Table 2.1 together with $\delta = 0.101957$ per year (Peterson et al., 1990).

Figure 10.4 illustrates the probability of a large outbreak of HIV when there are initially only a small number a of addicts and a small number b of needles infected. As $R_0 = 3.19 > 1$ for the parameters used this is given by the formula $1 - P_A^a P_B^b$. The figure shows clearly that the number of addicts initially infectious has a far greater influence on the probability of a major epidemic outbreak than the number of needles initially infectious. For example if four or more addicts (and no needles) are initially infectious then it is almost certain that a major epidemic outbreak will occur. In contrast if ten needles (and no addicts) are initially infectious then the probability of a major epidemic outbreak is only 0.02. This relationship is intuitively sensible as infected addicts remain in the population a lot longer than infected needles.

We now examine the impact of control measures such as needle exchange programs and improved needle cleaning in reducing the value of $1 - P_A^a P_B^b$. Figure 10.5 shows the

Figure 10.5: Impact of Control Measures on $1 - P_A$

$\lambda = 246.22$ per year

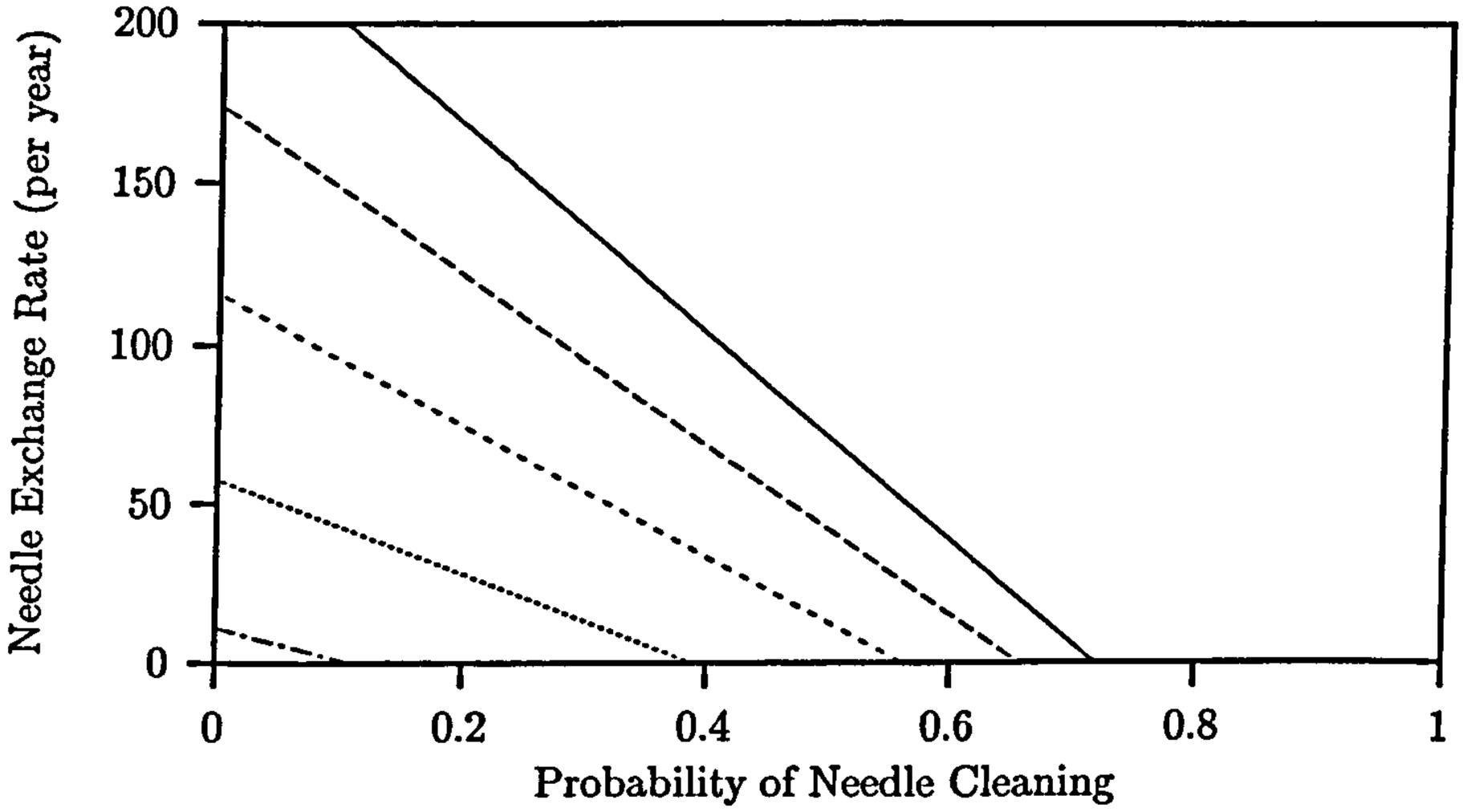


probability of a large outbreak of disease for various values of the needle exchange rate and the needle cleaning probability when a single addict and no needles are initially infectious. Kaplan (1995) estimates that once established a formal needle exchange program could increase the (average) needle exchange rate to 121.7 per year. We have previously estimated that without external intervention addicts clean needles prior to use approximately 64% of the time. The figure suggests that using interventions such as improved needle cleaning and needle exchange together could reduce the probability of a major epidemic outbreak to a very low level. However this would require that addicts both exchange needles on a regular basis (such as every few days) and that shared needles are cleaned prior to injection very often (for example over 80% of the time).

Figure 10.6 is similar to Figure 10.5 but where the shared injection rate, λ, has been lowered from 246.22 per year down to 100.0 per year. It is clear that interventions such as needle cleaning and needle exchange are now much more influential in reducing the likelihood of a major epidemic. For example if addicts clean needles prior to use only 64% of the time (as previously estimated) then the introduction of a needle exchange program on its own could reduce the probability of a major epidemic outbreak down to virtually zero, even if this program increased the needle turnover rate by only a small

Figure 10.6: Impact of Control Measures on $1 - P_A$

$\lambda = 100.0$ per year



amount. Such a reduction in the needle sharing rate may be possible if counselling and HIV testing are included as part of a needle exchange program. As discussed in Chapter 8 the motivation for such additions is that by being made aware of the risks involved addicts may reduce the rate at which they share needles and similarly infected addicts who are made aware of their positive HIV status (through taking an HIV test) would hopefully substantially reduce the rate at which they shared needles.

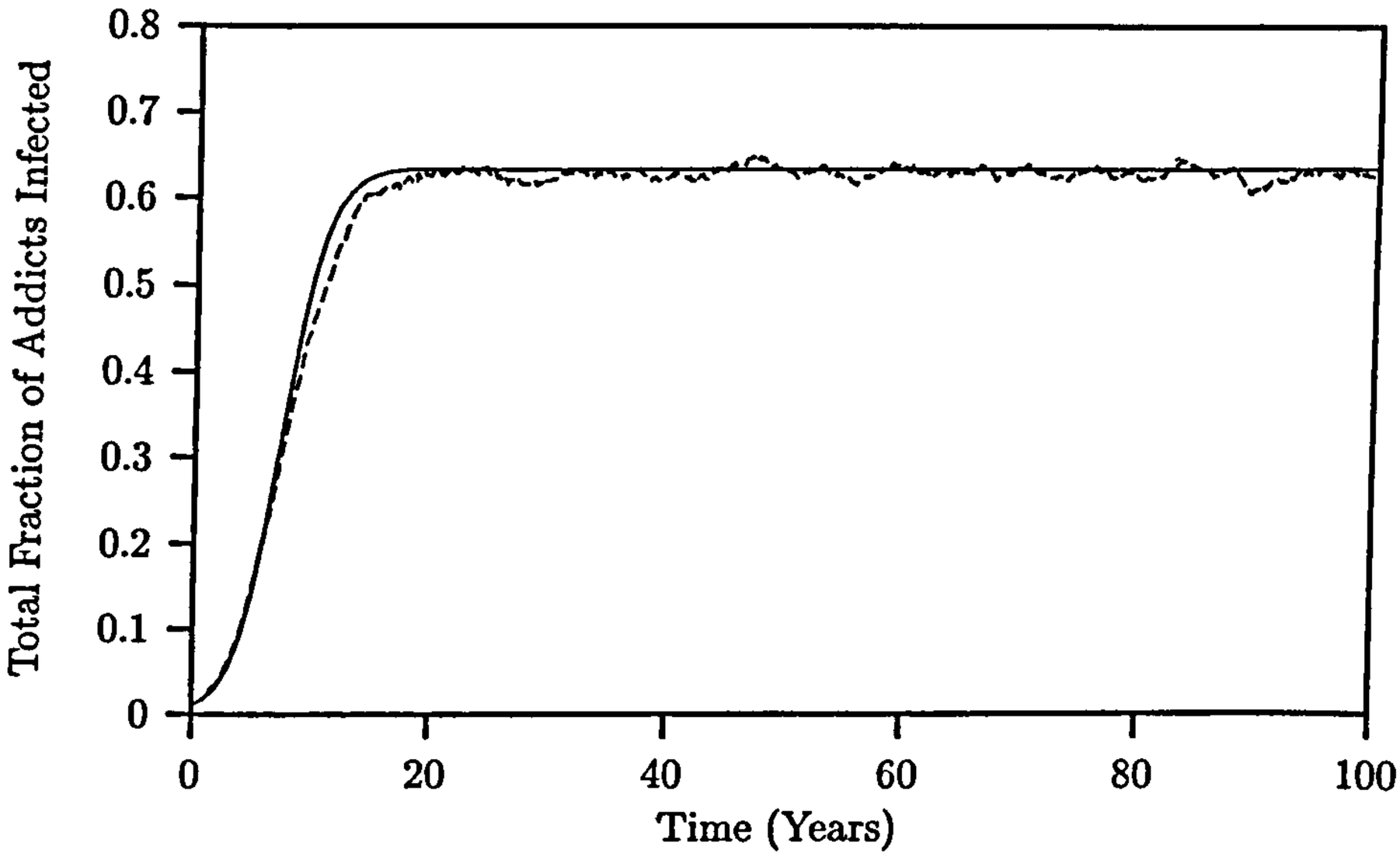
10.4 The Simple Model

Having discussed in detail the Kaplan and O'Keefe Model and its stochastic equivalent we now briefly take a similar look at the Simple Model. We examine the long term prevalence of disease in the stochastic and deterministic models and as usual use the parameter estimates from Table 3.1. Using an extension of the method in Theorem 10.1 it is possible to compute numerically the probability of extinction of disease in the Simple Model, and we conclude this section with a brief comparison of this probability and corresponding values for the Kaplan and O'Keefe Model.

Figure 10.7 shows the mean of the stochastic Simple Model conditional on the disease not dying out during the duration of the simulation period and the deterministic Simple Model. In the stochastic model it was assumed initially that one addict and

Figure 10.7: Simple Model Comparison

Conditional Mean, N=100, $R_0 > 1$

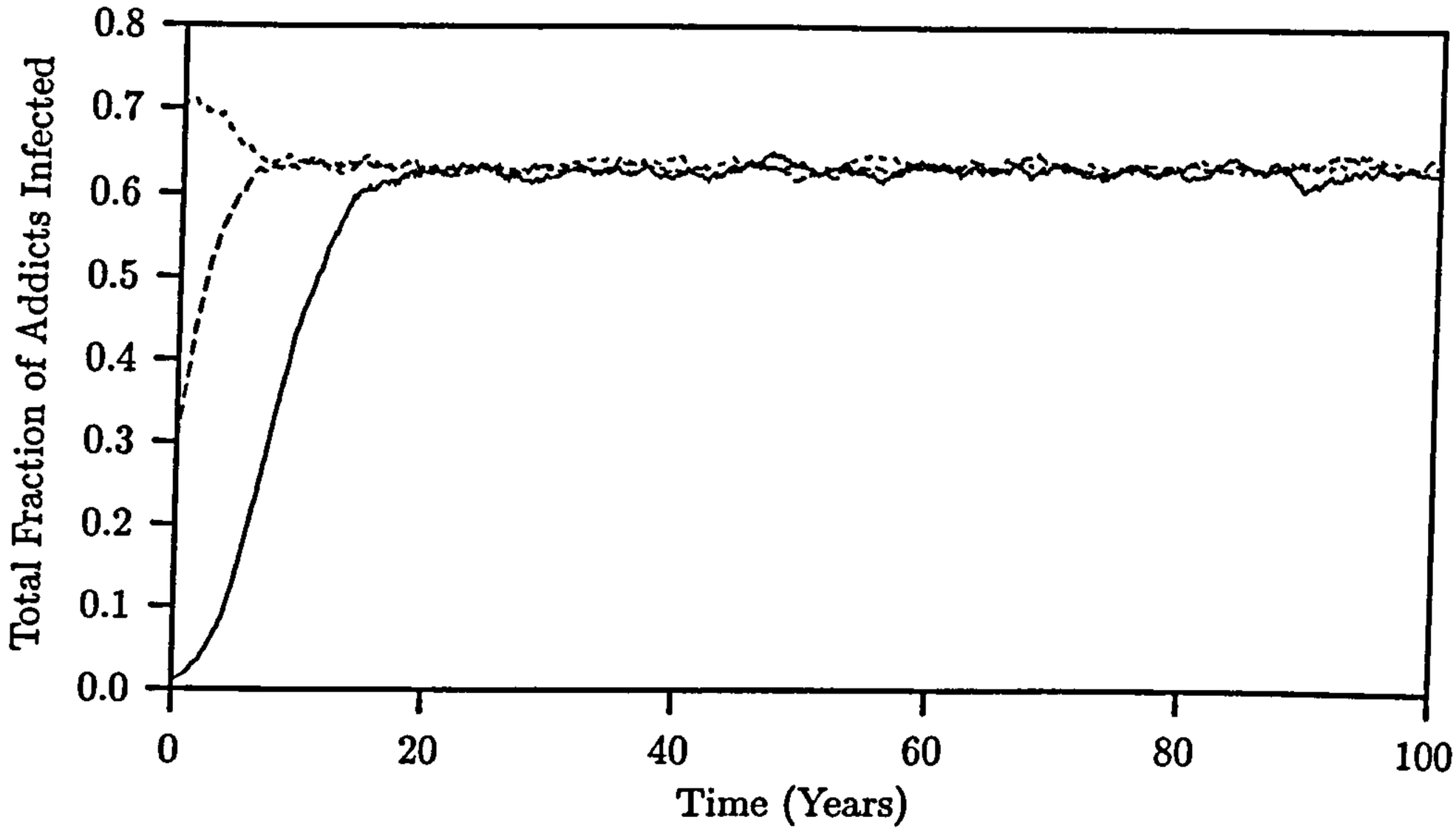


no needles were infectious where this single addict was assumed to be in stage one infectivity. In the deterministic model we assumed that 1.1% of the total addict population were infectious and no needles were infectious and again all initially infectious addicts were in stage one infectivity. It is clear from the figure that the long term prevalence of disease is very similar in both models. Figure 10.8 is similar to Figure 10.7 but shows the behaviour of the conditional mean of the stochastic Simple Model using three different initial conditions. The three simulations in this figure assume that initially 1, 27 and 64 addicts (among a total population of 91 addicts and 100 needles) are in stage one infectivity, this is approximately equivalent to initial prevalences of 1%, 30% and 70%. These simulations suggest that the quasi-equilibrium prevalence of disease in addicts is unaffected by the initial state of the disease (provided of course that disease is present in at least one addict or needle). This is again similar to the behaviour of the deterministic Simple Model.

We conclude our simulations of the stochastic Simple Model with an illustration of the effect of reducing the basic reproductive number to less than unity. Figure 10.9 shows a simulation of the unconditional mean of the Simple Model where initially one addict is in stage one infectivity and addicts clean needles prior to use with a probability of 0.64. We have assumed that after 25 years duration all addicts instantaneously

Figure 10.8: Simple Model Global Stability

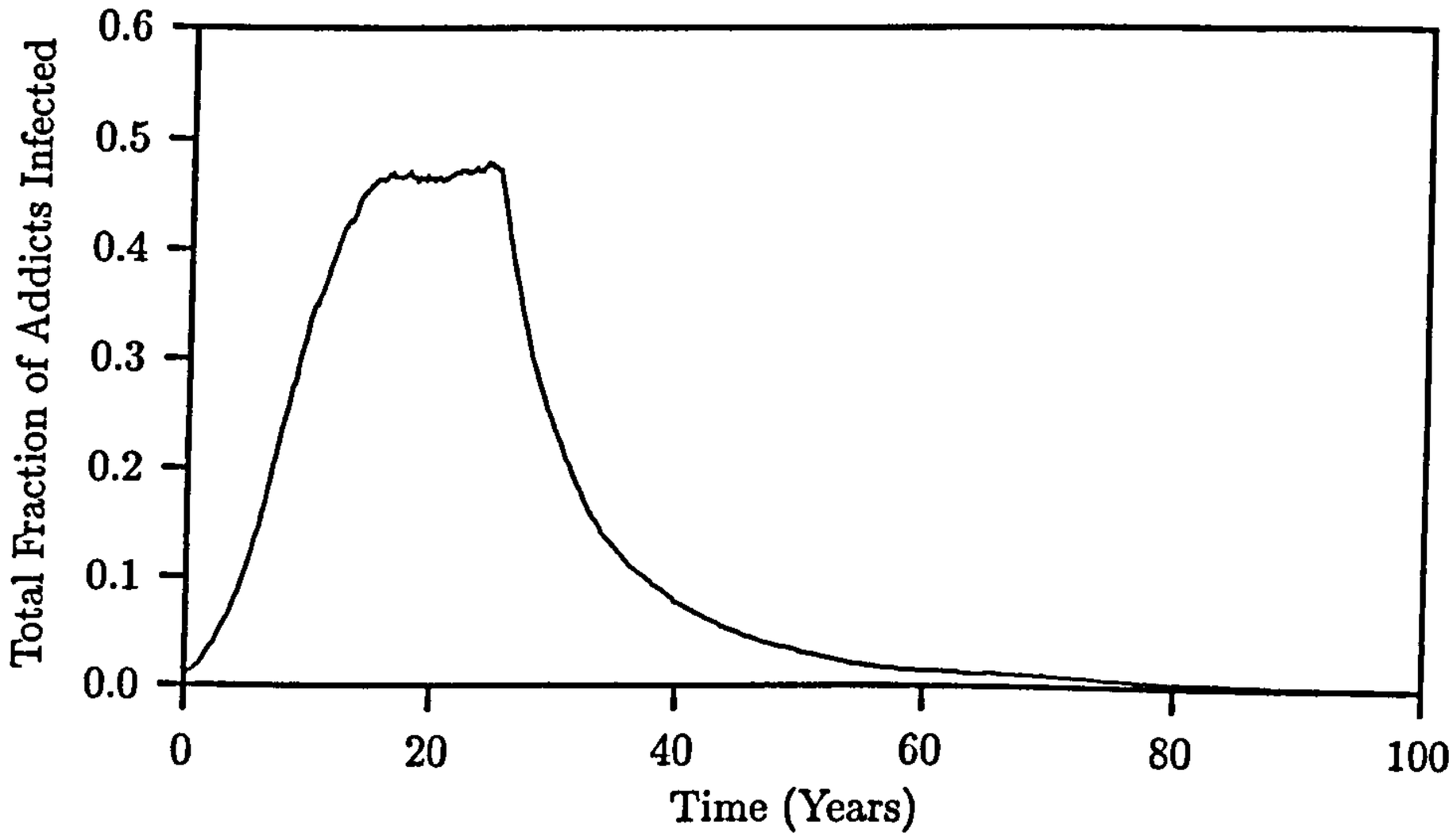
Conditional Mean, N=100, $R_0 > 1$



$\pi_1(0) \approx 0.01$ ——— $\pi_1(0) \approx 0.3$ ----- $\pi_1(0) \approx 0.7$ ------

Figure 10.9: Simple Model with Cleaning Intervention

Unconditional Mean, N=100



increase the probability of cleaning a needle prior to use from 0.64 to 0.89. This has the effect of reducing R_0 from 3.6 down to 0.9 at $t = 25$ years. We can see from the figure that the disease quickly reaches a quasi-endemic state before rapidly dying out due to the reduction of R_0 to less than the critical threshold value of $R_0 = 1$. Having illustrated that as with the Kaplan and O'Keefe Model, the stochastic Simple Model behaves in a similar fashion to its deterministic equivalent we now briefly examine the probability of extinction in the latter model in a similar manner to previously.

10.4.1 Probability of Extinction in the Simple Model

Using a branching process approximation similar to that used in Section 10.3.1, but this time based on the Simple Model, it is straightforward to prove the following theorem, analogous to Theorem 10.1 for the Kaplan and O'Keefe Model:

Theorem 10.2 *Suppose that there are initially $n + a$ addicts of whom a are infected and $m + b$ needles of which b are infected, n and m are large, and a and b are small. Then a large outbreak of HIV will occur with probability zero if $R_0 \leq 1$ and probability $1 - P_A^a P_B^b$ if $R_0 > 1$, where:*

$$q_i = \frac{\lambda}{\mu + \delta_i + \lambda}, \qquad \text{for } i = 1, 2, 3,$$

$$G_{X_1}(s) = \frac{1 - q_1}{1 - q_1 s},$$

$$G_{X_2}(s) = \frac{\mu}{\mu + \delta_1} + \frac{\delta_1}{\mu + \delta_1}\left(\frac{1 - q_2}{1 - q_2 s}\right)$$

and $$G_{X_3}(s) = 1 - \frac{\delta_1 \delta_2}{(\mu + \delta_1)(\mu + \delta_2)} + \frac{\delta_1 \delta_2}{(\mu + \delta_1)(\mu + \delta_2)}\left(\frac{1 - q_3}{1 - q_3 s}\right);$$

$$G_X(s) = G_{X_1}(s) G_{X_2}(s) G_{X_3}(s) \text{ and } G_Y(s) = 1 - \frac{c}{1 - p} + \frac{cs}{1 - ps};$$

$$G_Z(s) = G_X(G_Y(s));$$

$$G_{\tilde{Z}}(s) = G_Y(G_X(s));$$

$1 > P_A \geq 0$ is the unique root of $s = G_Z(s)$ in $[0, 1)$ and $1 > P_B \geq 0$ is the unique root of $s = G_{\tilde{Z}}(s)$ in $[0, 1)$.

Proof.

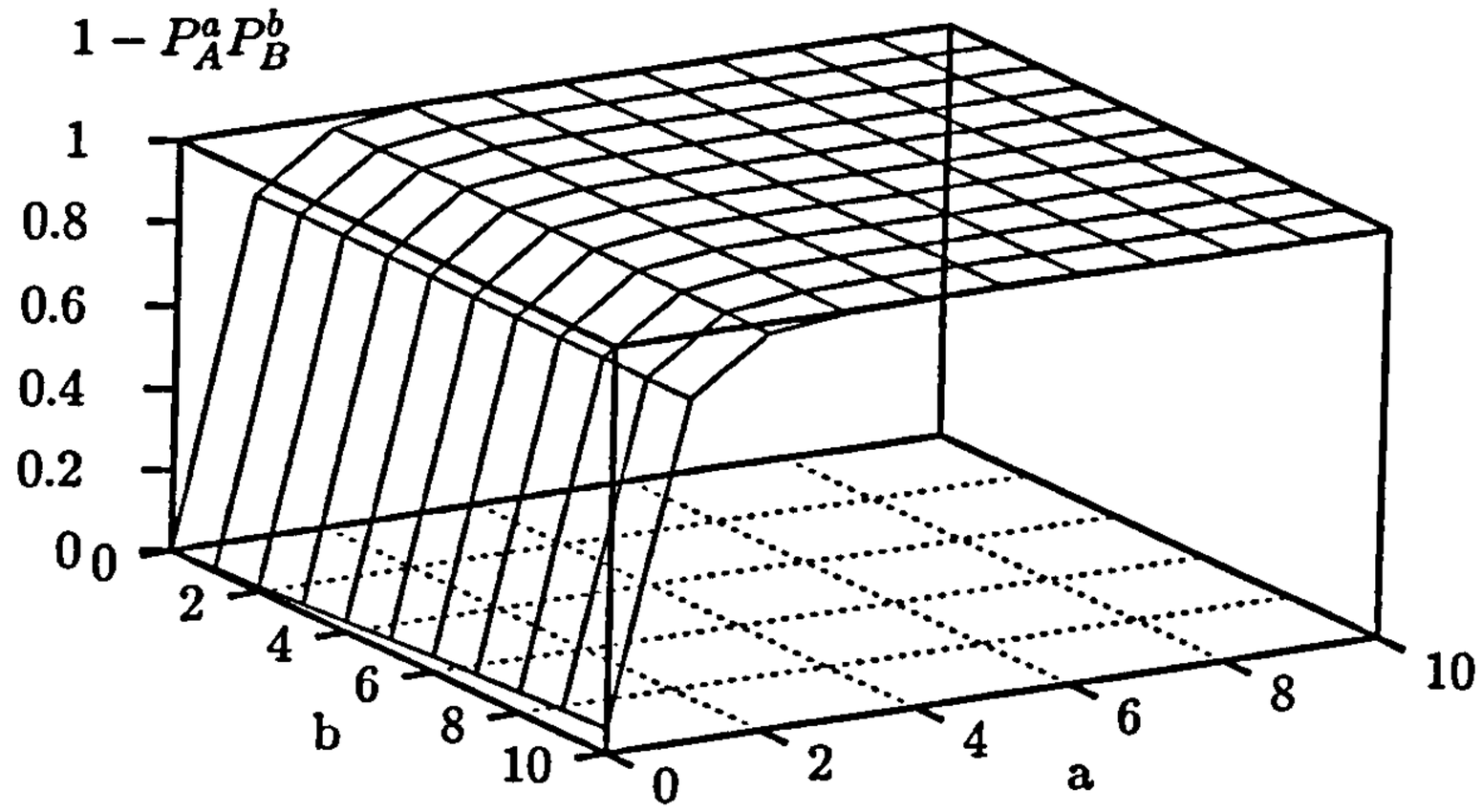
Let X denote the total number of needles infected by a single infectious addict during his or her entire infectious lifetime upon entering a population of n susceptible addicts and m uninfected needles. Therefore the random variable $X = X_1 + X_2 + X_3$, where for $i = 1, 2, 3$, X_i denotes the number of needles infected by a single infectious addict during his or her entire stage i infectious lifetime. Hence the p.g.f. of X is $G_X(s) = G_{X_1}(s)G_{X_2}(s)G_{X_3}(s)$ and the proof of this result follows very similarly to that of Theorem 10.1. If $R_0 > 1$ then the uniqueness of P_A and P_B follows by standard results in the theory of branching processes.•

Practical Implications

We find that the behaviour of the probability of a major epidemic outbreak in the Simple Model is qualitatively very similar to that of the Kaplan and O'Keefe Model. In particular it appears that as previously, while combining control measures such as needle exchange and needle cleaning can significantly reduce the probability of a major outbreak of disease, either measure is very much more effective if the needle sharing rate is also substantially reduced. Given our previous comparisons of single stage and three stage models in Chapter 7 it is of interest whether the probability of a major outbreak of disease is increased when we move to three stage infectivity in addicts.

Figure 10.10 is similar to Figure 10.4 but uses a value for the probability of HIV transmission in a single injection of $\alpha = 0.005342$ rather than $\alpha = 0.00601$. As discussed at length in Sections 7.2 and 7.5, using this value of α ensures that our three stage model and the Kaplan and O'Keefe Model are suitably calibrated in that we have that the cumulative viral load during an addict's infectious lifetime is the same in both models. All other parameters are as detailed in Table 2.1. From Figure 10.10 we again have that the number of initially infectious addicts has a very strong influence on the probability of a major outbreak and the number of initially infectious needles has a very limited effect. Moreover this figure also suggests that for any particular given combination of initially infectious addicts and needles the probability of a major outbreak is higher in the three stage model (under calibration).

Figure 10.10: $1 - P_A^a P_B^b$ for varying a and b (Three Stage Simple Model)



10.5 The Optimistic, Pessimistic and General Models

We now move on to comparing the Optimistic and Pessimistic Models with their stochastic equivalents. Figure 10.11 was constructed similarly to Figure 10.7 but features the deterministic Optimistic and Pessimistic Models and their stochastic equivalents. It is clear that the quasi-endemic equilibrium in the stochastic models is very similar to the endemic equilibrium suggested by the deterministic models. Simulations using a variety of different parameter estimates suggest that the deterministic models are good approximations of their more realistic stochastic equivalents. Figure 10.12 shows the impact of introducing a needle exchange program into our simulated population of addicts and needles where addicts and needles interact according to assumptions first in the Optimistic Model and then the Pessimistic Model. We expect from our previous discussion of the effect of needle exchange that increasing the needle exchange rate from $\tau = 15.53$ per year (the natural needle turnover rate) to $\tau = 121.7$ per year (the increased exchange rate resulting from the introduction of a formal exchange program) should produce a reasonably large decrease in the prevalence of disease. Figure 10.12 shows the conditional mean in the stochastic Optimistic and Pessimistic Models where we have initially assumed that 1.1% of all addicts are in stage one infectivity and needles are replaced with unused needles on average every 23.5 days (as estimated by Kaplan, 1995). We further assume that at $t = 50$ years a formal needle exchange

Figure 10.11: Optimistic and Pessimistic Model Comparison

Conditional Mean, N=100

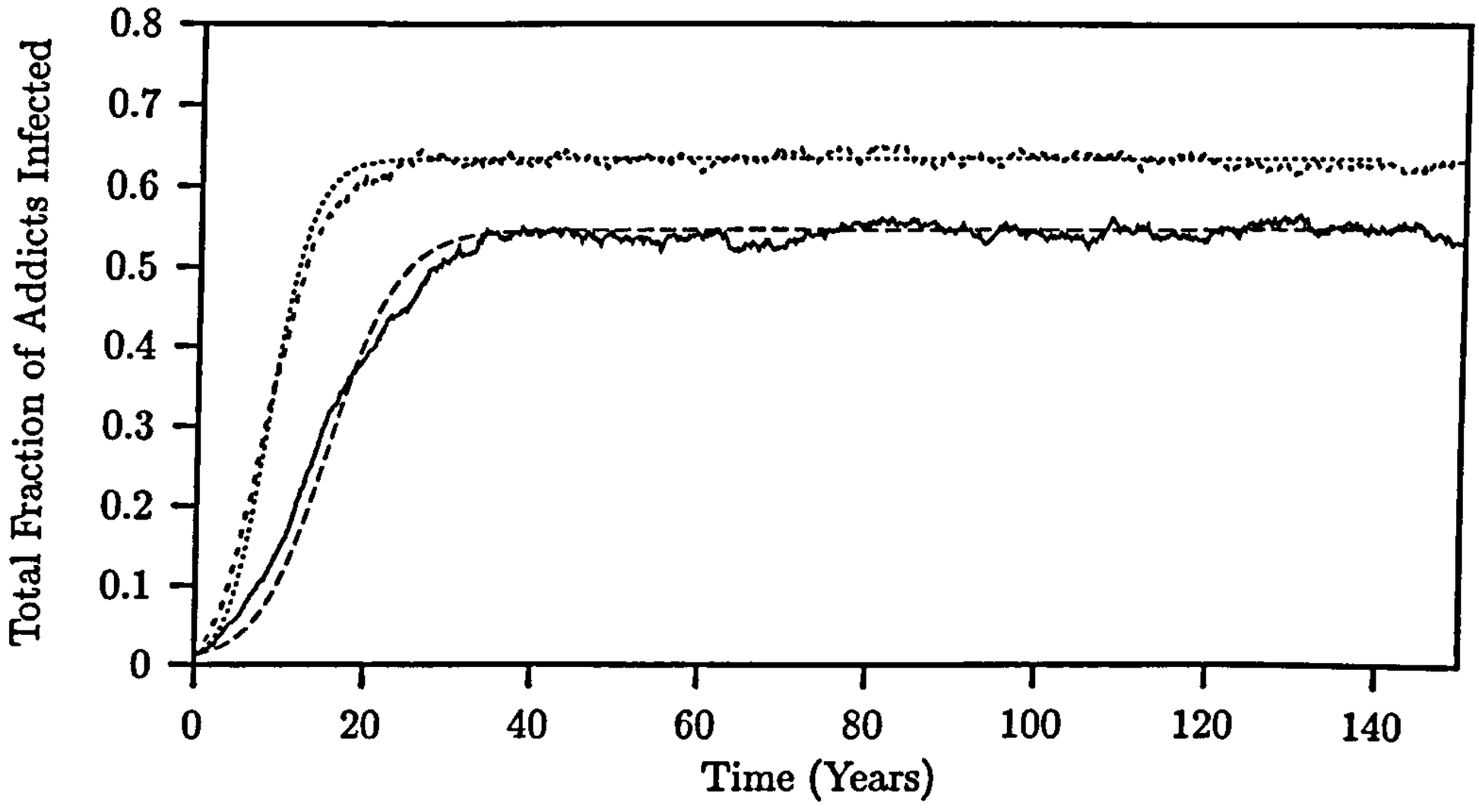


Figure 10.12: Optimistic and Pessimistic Models: Needle Exchange Intervention

Conditional Mean, N=100

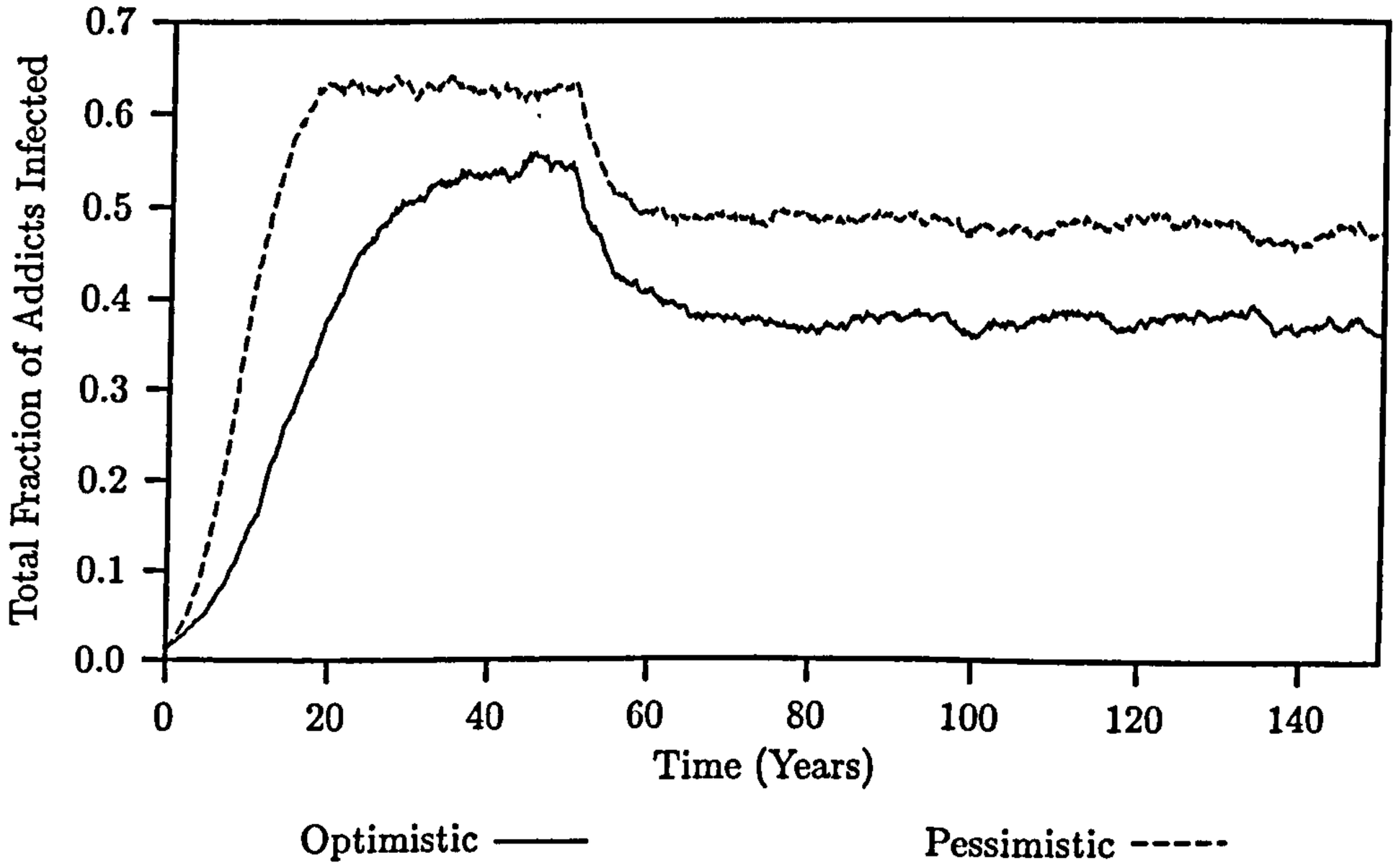
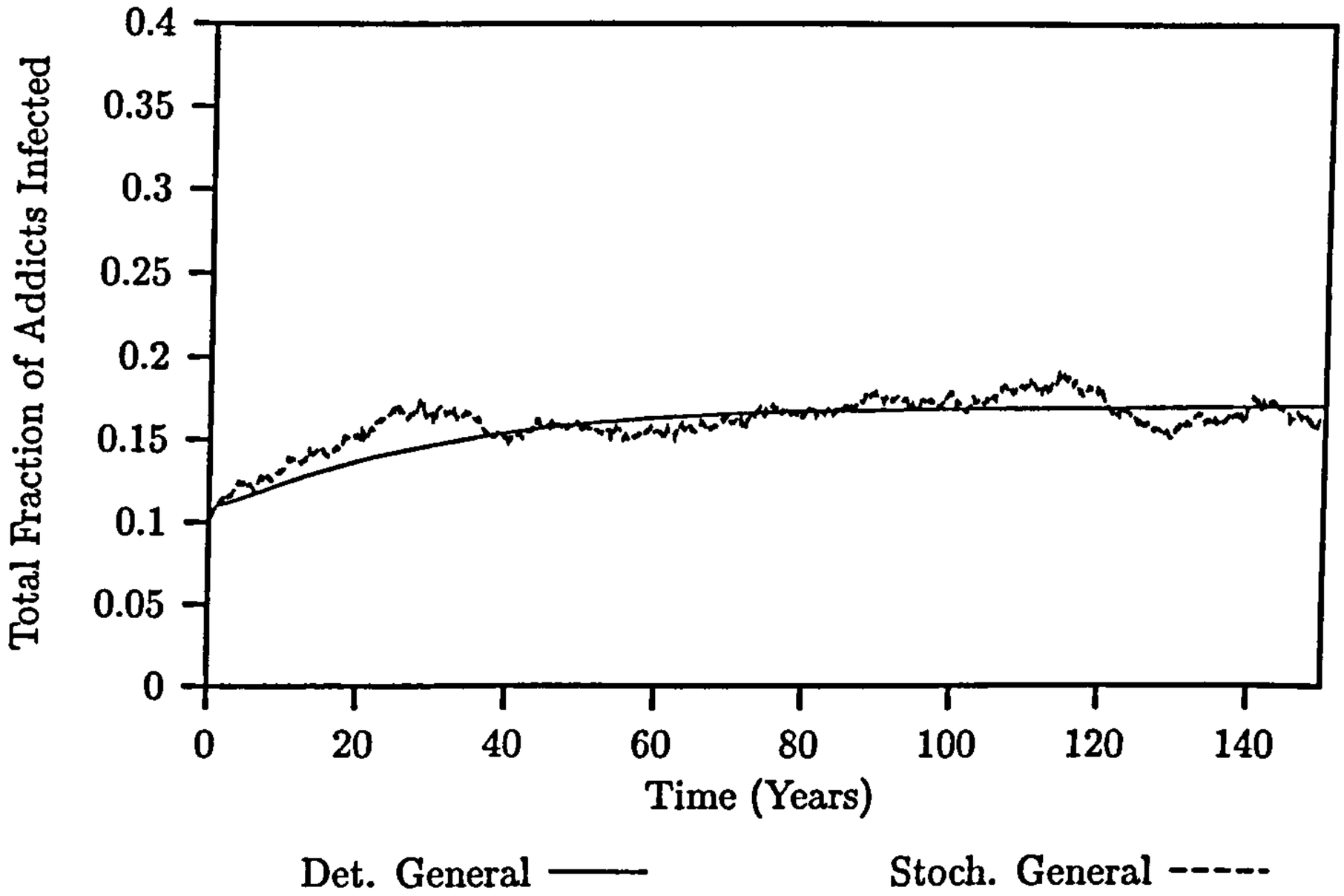


Figure 10.13: Equally Likely General Model

Conditional Mean, N=100



program has been established which decreases the average needle circulation time down to only 3 days. The simulations show a comparative level of decrease to that suggested by our deterministic models in Chapter 7.

We conclude our three stage infectivity model comparisons with a brief look at simulations of the (fully) General Model. Figure 10.13 shows the conditional mean for the General Model using the set of $p^{\star}_{ijk}$ terms denoted by B1 in Table 5.5 (the "equally likely" model). From Figure 5.5 in Chapter 5 we expect that this model will take a considerable amount of time to reach equilibrium. To reduce the computing time for this model we have assumed that initially nine addicts are in stage one infectivity and similarly that nine needles are in state one infectivity, the remaining addicts and needles being uninfectious, this means that the model should take less time to reach a steady state (since from the deterministic model we expect this model to tend to a prevalence in addicts of about 15% or equivalently about 14 addicts). From Figure 10.13 we can see that the conditional mean of the stochastic model is close to that of the equivalent deterministic model, however the conditional mean appears more variable than in some of our previous models (for example the Optimistic and Pessimistic Models). It was plausible that this could have been caused by the artificial starting conditions used in which the proportions of infected addicts and infected needles are out of phase with

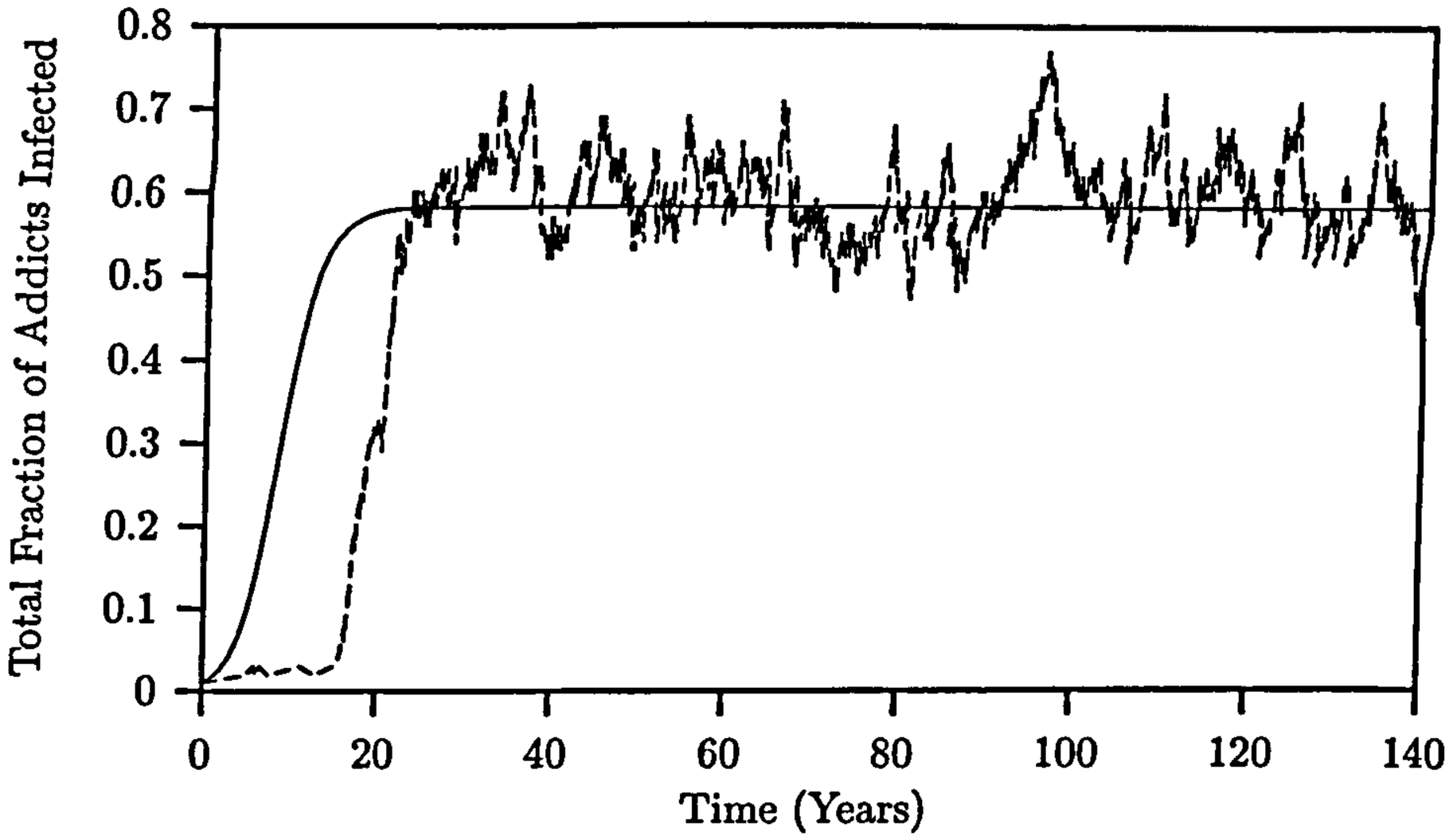
the equilibrium solution, causing initial large transient fluctuations in the prevalences of disease in addicts and needles. However individual simulations of this model did not appear to exhibit such initial fluctuations and the increased variability is more likely to have been due to the relatively low spread of disease suggested by this model together with the small total population size of 91 addicts and 100 needles used in our simulations. Using a much larger population size (and more simulation runs) would probably result in less variability, however this model is particularly computationally intensive taking 48 hours to produce the data shown in Figure 10.13. For this reason we do not illustrate simulations of this model for larger population sizes, but we found that while our simulations exhibit some additional variability they still suggest that the deterministic model is a good approximation of the equivalent stochastic process.

We have taken a brief look at comparing some of the three stage infectivity models which we have studied previously with their stochastic equivalents. We have found (perhaps not unexpectedly) that the criterion for an epidemic to take off or die out is the same, moreover our simulations suggest that given that disease take off actually occurs our deterministic models are very good approximations of their more realistic stochastic equivalents.

10.6 HIV Test Model

We now examine the behaviour of a stochastic model based on the same assumptions used to construct the HIV Test Model examined in Chapter 8. We examine this model in a similar fashion to those previously and again use a population size of $n = 91$ addicts and $m = 100$ needles. We are interested in any differences in long term behaviour and whether the deterministic model is a good approximation to the conditional expected value of the stochastic process. Figure 10.14 illustrates a single simulation of the stochastic HIV Test Model with $\lambda_1 = 250$ per year, $\lambda_2 = 150$ per year and where addicts are tested on average once every ten years ($\delta_t = 0.1$ per year) and where initially only a single addict was infectious and this addict injects at rate λ_1. Also shown in this figure is a numerical simulation of the deterministic HIV Test Model using the same parameter estimates where initially a fraction 0.011 of the addict population are infectious and unaware of their infectious status and all other addicts and needles are uninfectious. The figure shows that eventually the disease takes off and a considerable proportion of the addict population are infected. The prevalence of disease appears to reach a

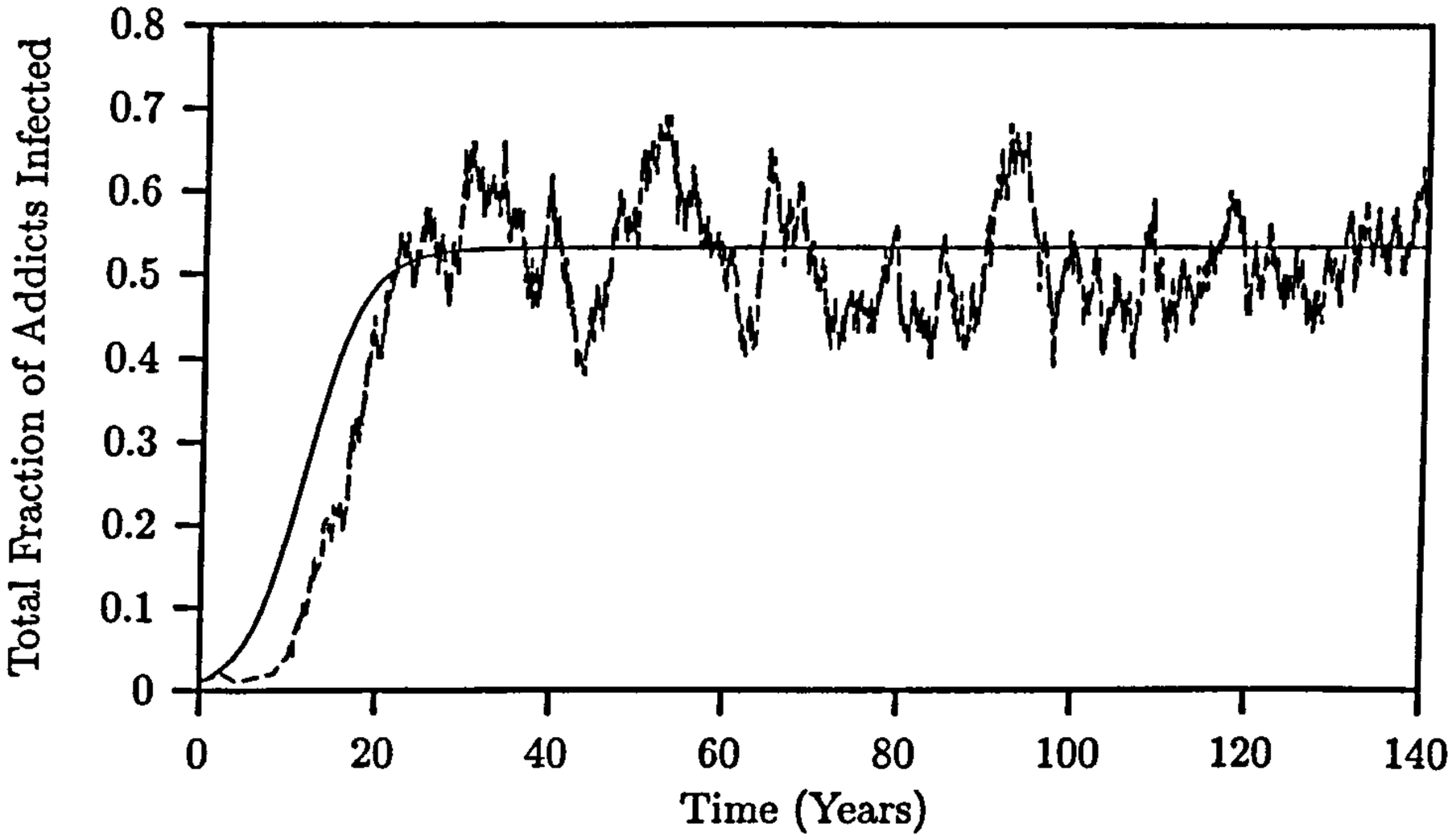
Figure 10.14: Single Realisation (HIV Test Rate=0.1 per year)



quasi-equilibrium state after about 30 years at which time the long term prevalence in the deterministic simulation seems to be an adequate description of the underlying level of disease (the level about which the stochastic process is fluctuating). A particularly interesting feature of the stochastic simulation is the initial behaviour of the disease. The disease remains almost dormant in the population for many years (between 15 and 16 years) before taking off very rapidly. It is important to note that this rapid take off is not caused by any change in behaviour of the population under study but is purely due to the random nature of disease transmission. While this behaviour only occurs in a small minority of the simulations of the HIV Test Model a similar phenomenon also occurs in a small minority of simulations of each of the stochastic models we have so far examined in this chapter. That is not to say that we expect this to occur in reality but it is clear that this behaviour could occur. For example it is conceivable (though admittedly unlikely) that a population in a particular country or region of the world may escape a major AIDS epidemic for many years even though a small amount of HIV was present in the population and then for no apparent reason a major epidemic occurs.

Figure 10.15 is similar to Figure 10.14 but the stochastic and deterministic models both assume that addicts are tested for HIV on average once a year. In this realisation we have again that the disease does not take off immediately but remains low for about ten years before taking off in earnest. Again this figure suggests that the deterministic

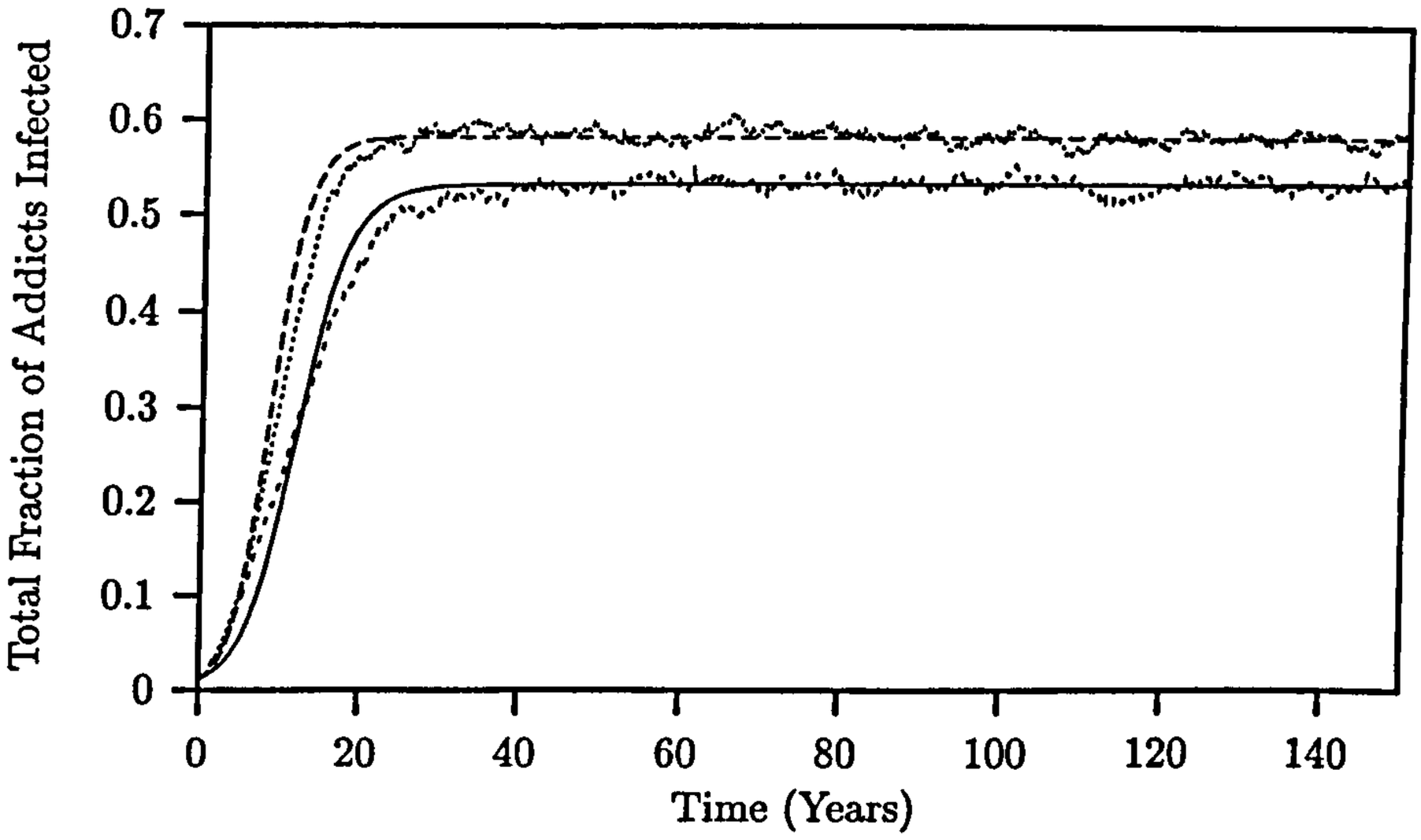
Figure 10.15: Single Realisation (HIV Test Rate=1.0 per year)



model is a good representation of the quasi-equilibrium state of the stochastic process. In addition this figure demonstrates clearly the large variability in the prevalence of HIV once the disease has passed the initial stage of rapid growth. For example at about 45 years the stochastic model reaches a prevalence of about 40% which for the next 15 years or so rapidly climbs to a prevalence of about 70%. It not difficult to see the problems that this level of variability can cause to public health policy decisions. If a steady increase in prevalence is observed then it is natural to assume that this sudden increase (or decrease) must have a specific cause, whether it be a new increased availability of cheap high quality heroin or *some other external factor*. However, as our simulations suggest, even an apparent sustained increase or decrease in the prevalence of disease may simply be due to the random nature of the epidemic and not the result of new behaviour or newly implemented control strategies.

We have examined several single simulations of the stochastic HIV Test Model, while these are useful and demonstrate the level of variability which may occur during an epidemic they are only one possible version of events. As in our previous models we now look at the conditional mean of the stochastic model as this represents our best estimate of how the disease may behave (given than an epidemic does in fact occur). Figure 10.16 shows the conditional mean based on 100 realisations of the stochastic process for the same parameter estimates and initial conditions used in Figures 10.14 and 10.15. We have also shown simulations of the deterministic HIV Test Model again using the

Figure 10.16: HIV Test Model: Deterministic v Stochastic Conditional Mean, N=100



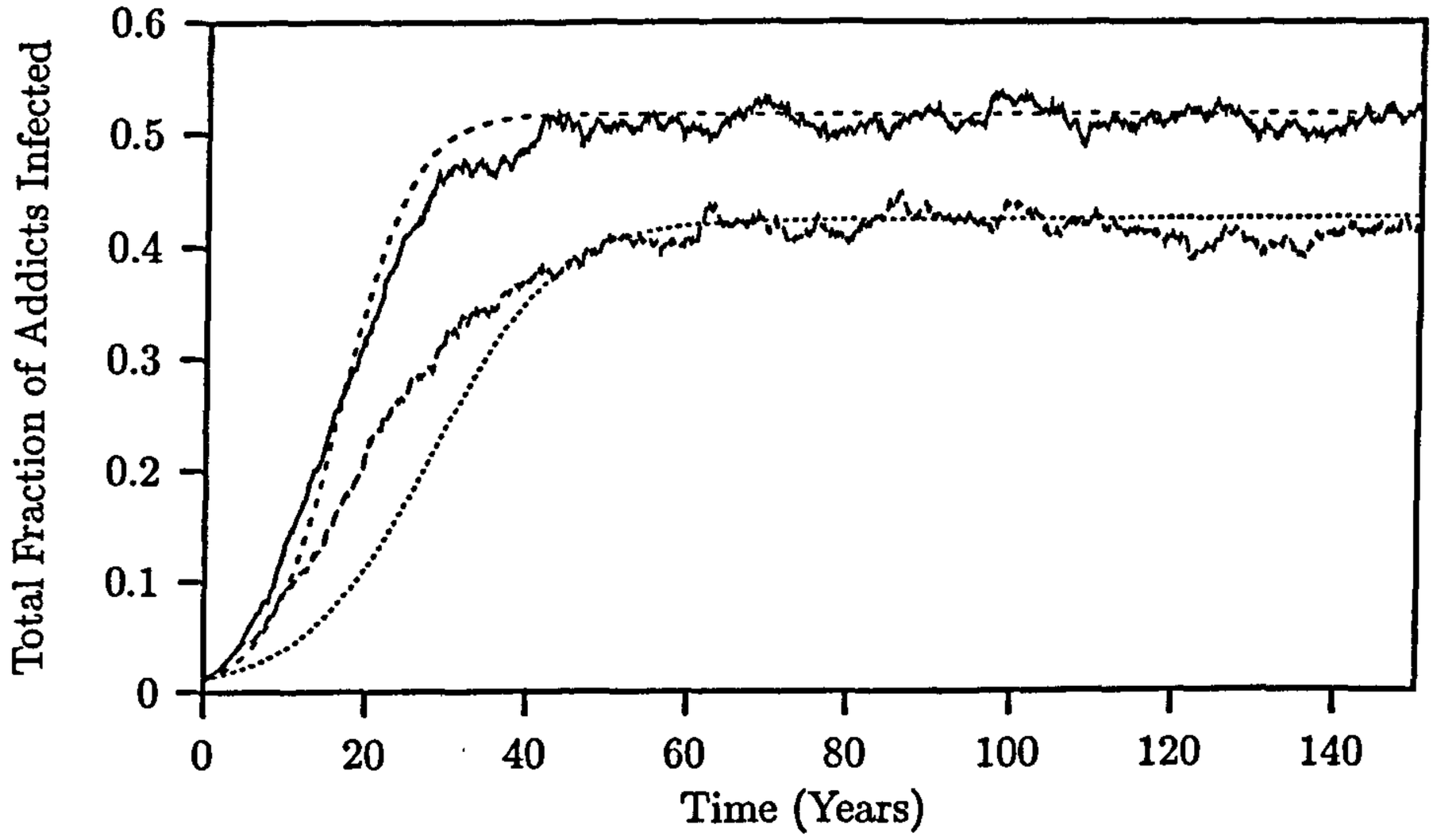
same parameter estimates as in these previous figures. It is clear that the deterministic model is a very good approximation to the conditional mean of the equivalent stochastic process. Simulations of the stochastic HIV Test Model suggest that $R_0 = 1$ is again the critical threshold point for this model between the disease dying out and taking off.

10.7 HIV Testing and Three Stage Infectivity

In a similar fashion to the previous sections we now examine the behaviour of stochastic equivalents to the Optimistic HIV Test Model and the Pessimistic HIV Test Model. As with our previous stochastic comparisons we find that the critical threshold (between the disease taking off or dying out), $R_0 = 1$, in the deterministic models is again the same for the stochastic models. Figure 10.17 shows a comparison of the Optimistic HIV Test Model with the equivalent stochastic process (where we use the conditional mean of this process as in our previous models). These simulations assume that prior to being tested for HIV addicts have a shared injection rate of 250 per year which drops to 150 per year upon receipt of a positive HIV test. We show two simulations of this model, in the first instance we have assumed that addicts are tested on average once every ten years. It is clear from the figure that the deterministic model is a good approximation to the conditional mean of this process for these parameter estimates.

Figure 10.17: Optimistic HIV Test Model ($\delta_t = 0.1, 1.0$)

Conditional Mean, N=100

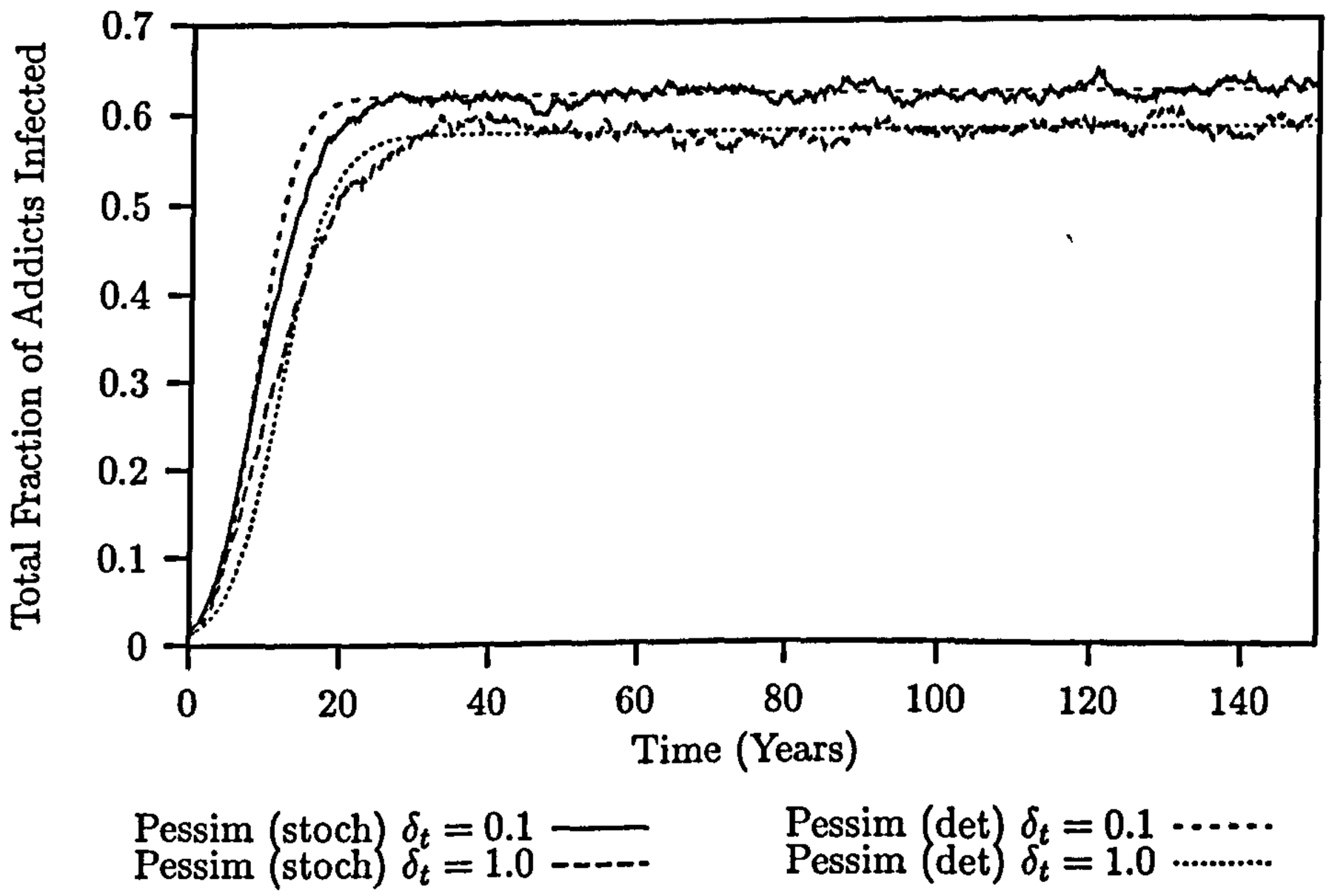


Optim (stoch) $\delta_t = 0.1$ ——— Optim (det) $\delta_t = 0.1$ ------

Optim (stoch) $\delta_t = 1.0$ - - - - - Optim (det) $\delta_t = 1.0$ .....

Figure 10.18: Pessimistic HIV Test Model ($\delta_t = 0.1, 1.0$)

Conditional Mean, N=100



Pessim (stoch) $\delta_t = 0.1$ ——— Pessim (det) $\delta_t = 0.1$ ------

Pessim (stoch) $\delta_t = 1.0$ - - - - - Pessim (det) $\delta_t = 1.0$ .....

The other simulation in the figure assumes that addicts are tested on average once a year, again it appears that the deterministic model is a good approximation to the long term prevalence level in the stochastic process.

Figure 10.18 is similar to Figure 10.17 but shows simulations of the Pessimistic HIV Test model. Again we have assumed that $\lambda_1 = 250$, $\lambda_2 = 150$ and $\delta_t = 0.1$ and 1.0. The figure suggests that the deterministic Pessimistic HIV Test Model is a good approximation to the conditional mean of the equivalent stochastic process. Other simulations (not illustrated) of the Pessimistic HIV Test Model for a variety of different parameter estimates suggest that this behaviour is typical.

10.8 Summary

In this chapter we have taken a brief look at comparing the deterministic models studied in detail in previous chapters, with stochastic models based on the same behavioural assumptions. We first outlined several of the fundamental differences between stochastic and deterministic models. We then examined the original Kaplan and O'Keefe Model and compared its behaviour with an equivalent stochastic process. We compared these models graphically using computer based simulations and then examined the probability of eventual extinction in this model using a branching process. We found that the basic reproductive number has a strong influence on the probability that the disease takes off, however this probability is not the reciprocal of the basic reproductive number as might have been expected from Whittle's Stochastic Threshold Theorem for the general epidemic model without vital dynamics.

Later we compared deterministic and stochastic equivalents of the Simple, Optimistic and Pessimistic Models together with a special case of the General Model. In each of these stochastic models we found that the critical threshold between the disease taking off and the disease dying out was the same as in the deterministic models, namely $R_0 = 1$. Of particular interest was whether the long term prevalence of disease suggested by our deterministic models was comparable with that of their more realistic stochastic alternatives. We found that the deterministic models were good approximations to the mean of the equivalent stochastic process conditional on the disease not dying out during the duration of the simulation period.

After comparing our deterministic three stage infectivity models we moved on to our models which included HIV testing. Simulations suggest that again the determin-

istic models are good approximations to their equivalent stochastic processes. We also examined several single realisations of the HIV Test Model. These suggest that the prevalence of disease can be highly variable once a quasi-equilibrium state has been reached. We commented that this high level of variability was true of all the stochastic models examined in this chapter and that this could pose considerable problems in evaluating the impact of control measures.

Chapter 11

Summary and Future Work

11.1 Summary

HIV and AIDS represents a serious health risk to populations of intravenous drug users. The sharing of injection equipment allows this virus to be transmitted among these populations with relative ease, compared to the population at large. A number of heterogeneous features have been shown to be important in affecting the spread of HIV through needle sharing. For example Greenhalgh (1996, 1997) has illustrated the importance of heterogeneity in the needle sharing rate and the efficiency of needle cleansing. The social networks in which addicts share needles have also been identified as potentially important features. Seitz and Müller (1994) have argued that variable infectivity can have a substantial effect on the spread of HIV among both drug users and the general population. In this thesis we have focused on examining the impact of variable infectivity specifically on the spread of HIV through needle sharing.

We first extended an established single stage infectivity model due to Kaplan and O'Keefe (1993) to allow addicts to progress through three distinct stages of infectivity. The basic infection mechanism in the Kaplan and O'Keefe Model was unaffected by this extension as we still maintained only a single type of infectious needle. This three stage model exhibited the same qualitative behaviour as the original, with the basic reproductive number as usual determining whether an epidemic takes off and the prevalence of disease reaches an endemic steady state or the disease dies out. Moreover we found that under suitable calibration, the long term quantitative behaviour of these models will always be the same. This is an intuitive result, although our extended model allows addicts to move through three stages of infectivity, we have assumed that

the infectivity of a needle used by an infectious addict is always the same irrespective of the infectivity of the addict.

If addicts progress through different stages of infectivity, it is obviously more realistic to suppose that the needles used by each of these types of addicts will contain different amounts of virus, and therefore represent different levels of risk to the next uninfectious addict to share any of these needles. Hence to model the transmission of HIV through needle sharing when addicts move through three stages of infectivity we should also incorporate three types of infectious needles, one for each type of infectious addict. This presents a problem since we are now required to specify how addicts and needles of different levels of infectivity interact with each other. In a single stage model we are required to determine the probability that a previously uninfectious needle becomes infectious after use by an infectious addict, and similarly the probability that a previously infectious needle becomes uninfectious after use by an uninfectious addict. Kaplan and O'Keefe estimate that the former probability equals unity and the latter probability equals zero. In a model with three types of infectious addicts and three types of infectious needles we have to specify similar probabilities, however there are now many more of these (64 in total). Moreover we are unaware of any data to assist in estimating many of these interaction probabilities.

Given the considerable problem of estimating how addicts and needles interact in a three stage infectivity framework, we derived models which represented intuitive upper and lower bounds of spread of disease. Our lower bound model was based on a generalisation of the case where an infectious needle is always left virus free after use by an uninfectious addict. Similarly our upper bound case was based on a generalisation of the case where an infectious needle is never left virus free by an uninfectious addict. We referred to our upper and lower bound models as the Pessimistic and Optimistic Models respectively. We demonstrated that again the basic reproductive number determines whether the disease dies out or reaches a unique endemic equilibrium. Using a suitable calibration method we compared the long term prevalence of disease in our upper and lower bound three stage models with equivalent upper and lower bound single stage models. We found that the lower bound models give rise to the same long term prevalence of disease in addicts and similarly for needles. The upper bound three stage model gives rise to a slightly higher long term prevalence of disease than in the equivalent single stage model. Hence from the comparisons so far it seems appropriate to conclude that three stage infectivity has little effect on the long term prevalence of

disease.

Estimating how addicts and needles of different levels of infectivity interact is difficult. While the Optimistic and Pessimistic Models are useful in assessing the effect of moving to three stage infectivity, they do assume that addicts and needles interact in very specific ways. To examine the impact of allowing addicts and needles to interact more broadly, we developed two generalised models of our simpler upper and lower bound models. These were the Restricted General Model and the General Model. These models contained many interaction parameters which we could not estimate precisely. However using analytical results it was possible to demonstrate that as in our previous models, if the basic reproductive number was less than unity then disease will die out in all addicts and all needles irrespective of how addicts and needles interact. Moreover in the Restricted General Model, the Pessimistic and Optimistic Models still represent special case upper and lower bounds. We examined a second general mixing model, the General Model which is a further generalisation of the first and removes a number of important restrictions regarding how addicts and needles interact. Specifically this model has a completely general addict-needle interaction structure and as such the Optimistic Model no longer represents a lower bound for the spread of disease, however the Pessimistic Model is still an upper bound.

To determine whether moving from single to three stage infectivity in general causes an increase in the long term prevalence of disease, we used various different calibration criteria to compare the spread of disease in our general mixing models with equivalent single stage models. We found that in general moving from single stage to three stage infectivity can result in either a significant increase in the prevalence of disease or a significant decrease in the prevalence of disease. Which of these occurs depends on the addict-needle interaction assumptions in our models. Two different calibration methods were used and both gave comparable results. Hence our models suggest that claiming, as Seitz and Müller do, that variable infectivity increases the prevalence of disease is not accurate, any increase (or decrease) additionally depends on how addicts and needles interact. As mentioned we are unaware of any data available to either justify or refute any particular set of addict-needle interaction assumptions.

In our three stage models we assumed that the size of the addict and needle populations remained constant throughout the course of an HIV epidemic. This is unrealistic as mortality from AIDS will reduce the number of addicts in the sharing, injecting population. In addition some recruitment of new susceptible addicts will undoubtably

occur, increasing the population size. It is highly unlikely that these two effects will always balance exactly to keep the population size constant. It was straightforward to extend our previous models to allow the population size to fluctuate in this manner. We found that this does not affect either the qualitative behaviour of our models, or the long term prevalence of disease among addicts and needles. However in this case it is the actual number of infectious addicts which is primarily of interest. In the general mixing models we found that different addict-needle interaction assumptions have less impact on the long term number of addicts infected than on the long term prevalence of disease among addicts. The simple reason for this is that in our variable population size models if the spread of disease is greater then so is AIDS mortality, and therefore while the prevalence of disease may be higher the difference between the actual number of addicts infected at equilibrium is reduced.

Once we had examined the impact of moving to three stage infectivity, we investigated different but related models which incorporated the testing of addicts for HIV. We first examined the behaviour of a model which was a more realistic extension of that discussed by Greenhalgh and Hay (1997). In this extended model we assumed that once infected, addicts were tested regularly for the presence of HIV. If an addict tested positive then he or she would share needles with a lower frequency than before. This model was more realistic than that examined by Greenhalgh and Hay, however on comparing these models we found that they had the same qualitative behaviour. Again the long term behaviour of our model was determined by the basic reproductive number. Moreover we showed that under suitable calibration these models were also quantitatively the same, in that they had the same endemic equilibrium solution.

Intuitively, the effectiveness of HIV testing as a control strategy should be closely linked to assumptions made relating to the variable infectivity of addicts during the long AIDS incubation period. For example suppose that addicts are highly infectious for only a few weeks after initial infection, and thereafter infectivity is very low until the development of full blown AIDS. In this case we would expect that HIV testing would not be a particularly effective control measure since by the time an addict has been tested, then he or she will probably be in the low infectivity state and as such be of relatively little importance in causing new infections. To investigate the impact that different relative infectivity assumptions had on the effectiveness of HIV testing we extended the three stage infectivity upper and lower bound Pessimistic and Optimistic Models discussed above to include HIV testing.

We found that the impact of HIV testing was significantly affected by different infectivity assumptions. As seems reasonable we found that the effect of HIV testing is greater for those assumptions in which more infectivity is concentrated in the latter part of the incubation period. We also found that for HIV testing to be at all effective, addicts need to be tested regularly and once aware of their infectious status greatly reduce the rate at which they share needles. Our models also suggest that addict-needle interaction assumptions have a significant impact on the effectiveness of HIV testing. Finally we suggested that the optimal use of HIV testing would be to combine it with a needle exchange program (as pioneered by Kaplan and O'Keefe). We believe that taken together these measures would provide a potentially very effective control strategy.

In Chapter 9 we moved away from our previous models and examined a method of sensitivity analysis for deterministic models. We discussed a method proposed by Blower and Dowlatabadi (1994) for determining the most influential parameters in complex models of disease transmission. We discussed a number of experimental designs and argued that the method chosen by Blower and Dowlatabadi had potential problems. Specifically they assume that the impact of any one model parameter is independent of the values of any other model parameter. This seems unlikely in a complex non-linear model. We argued that a better alternative was to repeat the method suggested by Blower and Dowlatabadi a number of times to get a fair idea of the importance of each model parameter, but that it was difficult to assess the number of repetitions required.

In Chapter 10 we briefly examined a number of stochastic alternatives to our previous deterministic models and computed an expression for the probability of extinction in the original Kaplan and O'Keefe Model. We found that the probability of extinction in the Kaplan and O'Keefe Model is not in general equal to $1/R_0$ as might have been expected from Whittle's Stochastic Threshold Theorem. Using extensive simulations we found that, perhaps surprisingly, given that a major outbreak of disease actually occurred all our deterministic models were very good approximations to the long term prevalence of disease suggested by the more realistic stochastic models. This is very encouraging as analysing the behaviour of stochastic models is much more difficult than for deterministic models. They are difficult analytically and require much more computational effort. This is particularly useful since the output from deterministic models is generally easier for non-specialists to understand. This is important when using models to justify policy decisions such as the introduction of HIV testing, needle

exchange and the promotion of needle cleaning.

11.2 Future Work

In this thesis we have examined in detail several aspects of the transmission of HIV among injecting drug users. This work has highlighted a number of interesting and possibly important areas of future work. Firstly it would be interesting to examine the effectiveness of control strategies such as needle exchange and HIV testing when only part of the population take up these measures. Taking this further and expanding on the work in Greenhalgh (1996), it would be interesting to see how the spread of disease is affected when the population consists of a small group of very high risk users, who practice little needle cleaning, frequently share needles and do not participate in any control measures. One interesting question to ask is how effective control measures would be in these circumstances when this small group (who will be mainly infectious) keep spreading disease among the remainder of the population. Another important way in which the work in this thesis could be extended would be to look more explicitly at the implications of needle sharing among groups of friends and social acquaintances as well as the sharing in shooting galleries on which Kaplan's original models were based.

Moving on to the larger picture it would interesting to see whether the claim by UNAIDS (1999), that intravenous drug use plays a fundamental role in the spread of HIV among the population at large is true. Or more specifically under what assumptions might this be true. For example using deterministic models of HIV transmission we could explore the mixing between drug using populations and heterosexual populations. This would determine how much and what kind of interaction between these two populations needs to occur for an epidemic to take place among the heterosexual population. This is an important practical issue as for many people in developed Western countries, such as those who are heterosexual and do not use intravenous drugs, HIV and AIDS is something which happens to others and these people perceive no risk to themselves. In worldwide terms the number of people newly infected with HIV is growing steadily, as is the number of individuals abusing intravenous drugs such as heroin. It seems plausible that if AIDS was to become the plague affecting all sectors of the population that was once predicted, then the spread of HIV among intravenous users and their interaction with the heterosexual population at large may be a crucial factor, particularly in developed countries and Second World countries such as the Russian

Federation, and it would be interesting to look at modelling this.

Bibliography

Abrams, E. J. (1987). *AIDS Bibliography 1986-1987.* U.S. Department of Health and Human Sciences.

Allain, J., Laurian, Y. and Paul, D. A. (1987). Long term evaluation of HIV antigen and antibodies to p24 and gp41 patients with hemophilia: potential clinical importance. *New England Journal of Medicine*, **317**, 1114-1120.

Anderson, R.M. (1988). The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity. *Journal of the Royal Statistical Society*, **A**, **151**, 66-93.

Anderson, R. M., Blythe, S. P., Gupta, S. and Konnings, E. (1989). The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behaviour. *Philosophical Transactions of the Royal Society of London*, **B**, **325**, 45-89.

Anderson, R. M., Gupta, S. and Ng, W. (1990). The significance of sexual partner contact networks for the transmission dynamics of HIV. *Journal of AIDS*, **3**, 417-429.

Anderson, R.M. and May, R.M. (1991). *Infectious Diseases of Humans: Dynamics and Control.* Oxford; Oxford University Press.

Anderson, R. M., May, R. M., Medley, G. F. and Johnson, A. M. (1986). A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. *IMA Journal of Mathematics Applied in Medicine and Biology*, **3**, 229-263.

Andersson, H. and Britton, T. (1998). Heterogeneity in epidemic models and its effect on the spread of infection. *Journal of Applied Probability*, **35**, 651-661.

Allard, R. (1990). A mathematical model to describe the risk of infection from sharing injection equipment. *Journal of Acquired Immune Deficiency Syndromes*, **3**, 1010-1016.

ANSWER (1995). *HIV and AIDS Surveillance in Scotland: Review of the Epidemic to December 1994.* Glasgow; SCIEH.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*, second edition. London; Griffin.

Barker, G. P., Berman, A. and Plemmons., R. J. (1978). Positive diagonal solutions to the Lyapunov equations. *Linear and Multilinear Algebra*, **5**, 249-256.

Barnard, M. and Frischer, M. (1995). Combining quantitative and qualitative approaches: Researching HIV related risk behaviours among drug injectors. *Addiction Research*, **2**, 351.

Bartlett, M. S. (1955). *Stochastic Processes.* Cambridge; Cambridge University Press.

Berman, A. and Hershkowitz, D. (1983). Matrix diagonal stability and its implications. *SIAM Journal of Algebraic and Discrete Methods*, **4**, 377-382.

Blower, S. M. and Dowlatabadi, H. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: An HIV model as example. *International Statistical Review*, **62**, 229-243.

Blower, S. M., Hartel, D., Dowlatabadi, H., Anderson, R. M. and May, R. M. (1991). Drugs, sex and HIV: A mathematical model for New York City. *Philosophical Transactions of the Royal Society of London*, **B**, **321**, 171-187.

Blythe, S. P. and Anderson, R. M. (1988a). Distributed incubation and infectious period models of the transmission of the human immunodeficiency virus (HIV). *IMA Journal of Mathematics Applied in Medicine and Biology*, **5**, 1-19.

Blythe, S. P. and Anderson, R. M. (1988b). Heterogeneous sexual activity models of HIV transmission in male homosexual populations. *IMA Journal of Mathematics Applied in Medicine and Biology*, **5**, 237-260.

Blythe, S. P., Castillo-Chavez, C., Palmer, J. S. and Cheng, M. (1991). Toward a unified theory of sexual mixing and pair formation. *Mathematical Biosciences*, **107**, 379-405.

BMBR (1987). AIDS advertising compaign, report on four surveys during the first year of advertising, 1986-1987. *British Market Research Bureau Limited.*

Capasso, V., Sicurallo, F. and Villa, M. (1995). Mathematical models for HIV transmission in groups of injecting drug users via shared drug equipment. *Journal of Biological Systems*, **3**, 747-758.

Castillo-Chavez, C., Cooke, K., Huang, W. and Levin S. A. (1989a). The role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS). In: *Mathematical Approaches to Problems in Resource Management and Epidemiology.* C. Castillo-Chavez, S. A. Levin and C. Shoemaker (Eds.) Lecture Notes in Biomathematics. **81**, 177-189. Berlin; Springer-Verlag.

Castillo-Chavez, C., Cooke, K., Huang, W. and Levin S. A. (1989b). On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS), Part 1. Single Population Models. *Journal of Mathematical Biology*, **27**, 373-398.

Castillo-Chavez, C., Cooke, K., Huang, W. and Levin S. A. (1989c). On the role of long incubation periods in the dynamics of acquired immunodeficiency syndrome (AIDS), Part 2. Multiple group models. In: *Mathematical and Statistical Approaches to AIDS Epidemiology.* C. Castillo-Chavez and S. A. Levin (Eds.) Lecture Notes in Biomathematics. **83**, 200-217. Berlin; Springer-Verlag.

Caulkins, J. P. and Kaplan, E. H. (1991). AIDS impact on the number of intravenous drug users. *Interfaces*, **21**, 50-63.

C. D. C. (1987). AIDS Weekly Surveillance Report: United States AIDS Program. Atlanta: Public Health Service, Center for Infectious Diseases, Center for Disease Control.

C. D. C. (1998). Morbidity and Mortality Weekly Report December 25, 1998. Atlanta: Public Health Service, Center for Infectious Diseases, Center for Disease Control.

Chow, S. N. and Hale, J. K. (1982). *Methods of Bifurcation Theory.* New York; Wiley.

Conover, W. J. (1980). *Practical Nonparametric Statistics.* New York; Wiley.

Cooke, K. L. and Yorke, J. A. (1973). Some equations modelling growth processes and gonorrhea epidemics. *Mathematical Biosciences*, **16**, 75-101.

Des Jarlais, D. C. and Freidman, S. R. (1987). HIV infection amongst intravenous drug users: epidemiology and risk reduction. *AIDS*, **1**, 67-76.

Des Jarlais, D. C., Freidman, S. R., Choopanaya, K., Vanichseni, S. and Ward, T. P. (1992). International epidemiology of HIV and AIDS amongst injecting drug users. *AIDS*, **6**, 1053-1068.

Des Jarlais, D. C., Hagan, H., Friedman, S. R., Friedman, P., Goldberg, D., Frischer, M., Green, S., Tunving, K., Ljungberg, B., Wodak, A., Ross, M., Purchase, D., Millson, M. E. and Myers, T. (1995). Maintaining low HIV seroprevalence in populations of injecting drug users. *Journal of the American Medical Association*, **274**, 1226-1231.

Diekmann, O., Heesterbeek, J. A. P. and Metz, J. A. J., (1990). On the definition and the computation of the basic reproductive ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, **28**, 365-382.

Dietz, K. (1988). The first epidemic model: a historical note on P. D. En'ko. *Australian Journal of Statistics*, **30A**, 56-65.

Dietz, K., Heesterbeek, J. A. P. and Tudor, D. W. (1993). The basic reproductive ratio for sexually-transmitted disease. 2. Effects of variable HIV infectivity. *Mathematical biosciences*, **117**, 35-47.

Dobyns, H. F. (1983). *Their Number Become Thinned: Native American Population Dynamics in Eastern North America*. Knoxville, USA; University of Tennessee Press.

Drucker, E. (1986). AIDS and addiction in New York City. *American Journal of Drug and Alcohol Abuse*, **12**, 165-181.

Fischl, M. A., Dickinson, G. M., Scott, G. M., Klimas, N., Fletcher, M. A. and Parks, W. (1987). Evaluation of heterosexual partners, children, and household contacts of adults with AIDS. *Journal of the American Medical Association*, **257**, 640-644.

Fisher, R. A. and Bennett, J. H. (1990) *Statistical methods, experimental design and scientific inferences: a re-issue of 'Statistical Methods for Research Workers', 'The Design of Experiments', and 'Statistical Methods and Scientific Inference'*. Oxford; Oxford University Press.

Friedland, G. H. and Klein, R. S. (1987). Transmission of the human immunodeficiency virus. *New England Journal of Medicine*, **18**, 1125-1135.

Frischer, M. (1997). Mortality and survival among a cohort of drug injectors in Glasgow, 1982-1994. *Addiction*, **92**, 419-427.

Gani, J. (1990). Epidemic modeling and simulation. *Mathematics and Computers in Simulation*, **32**, 3-12.

Gani, J. and Yakowitz, S. (1993). Modeling the spread of HIV among intravenous drug users. *IMA Journal of Mathematics Applied in Medicine and Biology*, **10**, 51-65.

Ginzburg, H. M. (1984). Intravenous drug users and the acquired immune deficiency syndrome. *American Journal of Public Health*, **99**, 206-212.

Goldberg, D., Frischer, M., Green, S. T., Taylor, A. and McKeganey, N. (1995). Probability of HIV transmission among *injecting drug users in* Glasgow. Unpublished manuscript.

Green, S. T., Taylor, A., Frischer, M. and Goldberg, D., J. (1993). Frontloading (halfing) among Glasgow drug injectors as a continuing risk behaviour for HIV transmission. *Addiction*, **88**, 1581-1582.

Greenhalgh, D. (1996). Effects of heterogeneity on the spread of HIV/AIDS among intravenous drug users in shooting galleries. *Mathematical Biosciences*, **136**, 141-186.

Greenhalgh, D. (1997) Bounds on endemic disease levels, risks and basic reproductive numbers in heterogeneous models for HIV/AIDS amongst injecting drug users. *Journal of Biological Systems*, **5**, 151-180.

Greenhalgh, D. and Hay, G. (1997). Mathematical modelling of the spread of HIV/AIDS amongst injecting drug users. *IMA Journal of Mathematics Applied in Medicine and Biology*, **14**, 11-38.

Grund, J., Kaplan, C. D., Adriaans, N. F. P., Blanken, P. and Huisman, J. (1990). The limitations of the concept of needle sharing: The practice of frontloading. *AIDS*, **4**, 815-23.

Hale, J. K. (1969). *Ordinary Differential Equations.* New York; Wiley.

Hay, G. (1999). *Modelling the Spread of HIV/AIDS amongst Injecting Drug Users.* PhD thesis, Glasgow, UK; University of Strathclyde.

Hethcote, H. W. and Yorke, J. A. (1984). *Gonorrhea: transmission dynamics and control. Lecture Notes in Biomathematics*, **56**, 1-105.

Heesterbeek, J. A. P. and Dietz, K. (1996). The concept of R_0 in epidemic theory. *Statistica Neerlandica*, **50**, 89-110.

Huang, W. (1989). PhD. dissertation, Claremont, USA; Claremont Graduate School.

Huang, W., Cooke, K. and Castillo-Chavez, C. (1992). Stability and bifurcation for a multiple group model for the dynamics of HIV/AIDS. *SIAM Journal of Applied Mathematics*, **52**, 835-854.

Hunter, G. M., Donoghoe, M. C., Stimson, G. V., Rhodes, T. and Chalmers, C. P. (1995). Changes in the injecting risk behaviour of injecting drug-users in London, 1990-1993. *AIDS*, **9**, 493-501.

Hyman, J. M., Li, J. and Stanley, E. A. (1999). The differential infectivity and staged progression models for the transmission of HIV. *Mathematical Biosciences*, **155**, 77-109.

Iannelli, M., Milner, F. A., Pugliese, A. and Gonzo, M. (1997). The HIV/AIDS epidemics among drug injectors: A study of contact structure through a mathematical model. *Mathematical Biosciences*, **139**, 25-58.

Isham, V. (1988). Mathematical modelling of the transmission dynamics of HIV infection and AIDS: A review. *Journal of the Royal Statistical Society*, **A**, **151**, 5-30.

Johnson, C. R. (1974). Sufficient conditions for D-stability. *Journal of Economic Theory*, **9**, 53-62.

Jordan, D. W. and Smith, P. (1987). *Nonlinear Ordinary Differential Equations*, second edition. Oxford; Clarendon Press.

Kanki, P. J., Alroy, J. and Essex, M. (1985). Isolation of a T-lymphotropic retrovirus type-4 and the human immunodeficiency virus in West Africa. *Science*, **236**, 827-831.

Kaplan, E. H. (1989a). Needles that kill: Modeling human immunodeficiency virus transmission via shared drug injection equipment in shooting galleries. *Reviews of Infectious Diseases*, **11**, 289-298.

Kaplan, E. H. (1989b). Can bad models suggest good policies? Sexual mixing and the AIDS epidemic. *Journal of Sex Research*, **26**, 301-314.

Kaplan, E. H. (1996). Private communication.

Kaplan, E. H. and Heimer, R. (1992a) HIV prevalence among intravenous drug users: Model based estimates from New Haven's legal needle exchange. *Journal of Acquired Immune Deficiency Syndromes*, **5**, 163-169.

Kaplan, E. H. and Heimer, R. (1992b) A model based estimate of HIV infectivity via needle sharing. *Journal of Acquired Deficiency Syndromes*, **5**, 1116-1118.

Kaplan, E. H. and Lee, Y. S. (1990). How bad can it get? Bounding worst case endemic heterogeneous mixing models of HIV/AIDS. *Mathematical Biosciences*, **99**, 157-180.

Kaplan, E. H. and O'Keefe, E. (1993) Let the needles do the talking! Evaluating the New Haven needle exchange. *Interfaces*, **23**, 7-26.

Kaplan, E. H. (1994) A method for evaluating needle exchange programs. *Statistics in Medicine*, **13**, 2179-2187.

Kaplan, E. H. (1995) Probability models of needle exchange. *Operations Research*, **43**, 558-568.

Kermack, W. O. and McKendrick, A. G. (1927-33). Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society*, **A**, **115**, 700-721. (Part I, 1927), **A**, **138**, 55-83. (Part II, 1931), **A**, **141**, 94-122. (Part III, 1933).

Koopman, J. S., Jacquez, J. A., Welch, G. W., Simon, C. P., Foxman, B., Pollock, S. M., Brath-Jones, D., Adams A. L. and Lange K. (1997). The role of early HIV infection in the spread of HIV through populations, *Journal of Acquired Immune Deficiency Syndromes*, **14**, 249-258.

Kriess, J. K., Kitchen, L. W., Prince, H. E., Kasper, C. K. and Essex, M. (1985). Antibody to human T-lymphototropic virus type III in wives of hemophiliacs: evidence for heterosexual transmission. *Annual of International Medicine*, **102**, 623-626.

Kretzschmar, M. and Wiessing, L. G. (1998). Modelling the spread of HIV in social networks of injecting drug users. *AIDS*, **12**, 801-811.

Lajmanovich, A. and Yorke, J. A. (1976). A deterministic model for gonorrhea in a nonhomogeneous population. *Mathematical Biosciences*, **28**, 221-236.

La Salle, J. P. (1976). *The Stability of Dynamical Systems: Regional Conference Series in Applied Mathematics.* Philadelphia; SIAM.

Leentvaar-Kuijpers, A., Dekker, M. M., Countinho, R. A., Dekker, E. E. and Keeman, J. N. (1990). Needlestick injuries, surgeons and HIV risks. *Lancet*, **335**, 546-547.

Lin, X. (1991). On the uniqueness of endemic equilibria of an HIV/AIDS transmission model for a hetergeneous population. *Journal of Mathematical Biology*, **29**, 779-790.

Longini, I. M., Clark, W. S., Byers, R. H., Ward, J. W., Darrow, W. W., Lemp, G. F. and Hethcote, H. W. (1989). Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine*, **8**, 831-843.

Longini, I. M., Clark, W. S., Gardner, L. I. and Brundage, J. F. (1991). The dynamics of $CD4^+$ T-lymphocycte decline in HIV-infected individuals: A Markov modelling approach. *Journal of Acquired Immune Deficiency Syndromes*, **4**, 1141-1147.

Lyapunov, A. (1892). Problemes général de la stabilité du mouvement, *Communications of the Mathematical Society of Kharkov,* republished in *Annals of Mathematical Studies,* **18**. Princeton; Princeton University Press 1947.

Macdonald, G. (1950). The analysis of infection rates in diseases in which superinfection occurs. *Tropical Diseases Bulletin*, **47**, 907-915.

Macdonald, G. (1952). The analysis of equilibrium in malaria. *Tropical Diseases Bulletin*, **49**, 813-829.

Macdonald, G. (1953). The analysis of malaria epidemics. *Tropical Diseases Bulletin*, **50**, 871-889.

Macdonald, G. (1957) *The Epidemiology and Control of Malaria.* London; Oxford University Press.

MAP (1998). *Provisional Report from MAP (Monitoring the AIDS Pandemic): The Status and Trends of the HIV/AIDS Epidemics in the World.* Available on the Internet from UNAIDS at www.unaids.org.

Massad, E., Countinho, F. A. B., Yang, H. M., Decarvalho, H. B., Mesquita, F. and Burattini, M. (1994). The basic reproductive ratio of HIV among intravenous drug users. *Mathematical Biosciences*, **123**, 227-247.

May, R.M. (1973). *Stability and Complexity in Model Ecosystems.* Princeton: Princeton University Press.

May, R.M., and Andersón, R.M. (1987). Transmission dynamics of HIV infection. *Nature*, **326**, 137-142

McKeganey, N. and Barnard, M. (1992). *AIDS, Drugs and Sexual Risk: Lives in the Balance.* Buckingham; Open University Press.

Mollison, D. (1991) Dependence of epidemic and population velocities on basic parameters. *Mathematical Biosciences*, **107**, 255-287.

Nikiforuk, A. (1992). *The Fourth Horseman.* London; Fourth Estate Limited.

Nold, A. (1980). Heterogeneity in disease transmission modelling. *Mathematical Biosciences*, **52**, 227-240.

Osmond, D. (1990). Sexual transmission of HIV infection: overview. *The AIDS knowledge base*, Cohen, P.T., Sande, M. A. and Volberding, P. A. (Eds.) Waltham, MA; Medical Publishing Group.

Page, J. B. (1990). Shooting scenarios and risk of HIV-1 infection. *American Behavioural Science*, **33**, 478-490.

Peterman, T. A., Stoneburner, R. L. and Allen, J. R. (1986). Risk to HTLV-III/LAV transmission to household contacts of persons with transfusion-associated HTLV-III/LAV infection. *Conference Proceedings of the Second International Conference on AIDS, Paris*, 102-113.

Peterson, D., Willard, K., Altmann, M., Gatewood, L. and Davidson, G. (1990). Monte-Carlo simulation of HIV infection in an intravenous drug user community. *Journal of Acquired Immune Deficiency Syndromes*, **3**, 1086-1095.

Redfield, R. R. and Burke, D. S. (1988). HIV infection: the clinical picture. *Scientific American*, **259**, 90-99.

Redfield, R. R., Markham, P. D. and Salahuddin, S. Z. (1985). Frequent transmission of HTLV-III among spouses of patients with AIDS-related complex and AIDS. *Journal of the American Medical Association*, **253**, 1571-1573.

Renshaw, E. (1991) *Modelling Biological Populations in Space and Time.* Cambridge Universty Press; Cambridge.

Ross, R. (1908). *Report on the Prevention of Malaria in Mauritius.* London; Murray.

Ross, R. (1911). *The Prevention of Malaria*, (2nd Edn.). London; Murray.

Ross, R. (1915). Some *a priori* pathometric equations. *British Medical Journal*, **1**, 546-547.

Ross, R. (1916). An application of the theory of probabilities to the study of *a priori* pathometry, I. *Proceedings of the Royal Society*, **A**, **92**, 204-230.

Ross, R. (1917). An application of the theory of probabilities to the study of *a priori* pathometry, II. *Proceedings of the Royal Society*, **A**, **93**, 225-240.

Samuels, J. F., Vlahov, D., Anthony, J. C. and Chaisson, R. E. (1992). Measurement of HIV risk behaviours amongst intravenous drug users. *British Journal of Addiction*, **87**, 417-428.

Sattenspiel, L. and Castillo-Chavez, C. (1990). Environmental context, social interactions, and the spread of HIV. *American Journal of Human Biology*, **2**, 341-366.

Sattenspiel, L., Koopman, J., Simon, S. and Jacquez, J. A. (1990). The effects of population structure on the spread of HIV infection. *American Journal of Physical Anthropology*, **82**, 421-429.

Schwager, S. J., Castillo-Chavez, C. and Hethcote, H. W. (1989). Statistical and Mathematical approaches in HIV/AIDS modelling: a review. In: *Mathematical and Statistical Approaches to AIDS Epidemiology*. C. Castillo-Chavez and S. A. Levin (Eds.). Lecture Notes in Biomathematics, **83**, 2-35 Berlin; Springer-Verlag.

Seitz, S. T. and Müller, G. E. (1994). Viral load and sexual risk. Epidemiological and policy implications for HIV/AIDS. In: *Modeling the AIDS Epidemic: Planning,*

Policy and Prediction, Kaplan, E. H. and Brandeau, M. L. (Eds.), 461-480. New York; Raven Press.

Stanley, E. A. and Hyman, J. M. (1988). Using mathematical models to understand the AIDS epidemic. *Mathematical Biosciences*, **90**, 415-473.

Stimson, G. V., Alldritt, L., Dolan, K. and Donoghoe, M. (1988). Injecting equipment exchange schemes: A preliminary report on research. Monitoring Research Group Technical Report. London: University of London Goldsmith's College.

Stimson, G. V., Des Jarlais, D. C. and Ball, A. (1997). *Drug Injecting and HIV Infection*. London; UCL Press.

Tan, W. Y. and Tang, S. C. (1993). A stochastic model of the HIV epidemic involving both sexual contact and IV drug use. *Mathematical and Computer Modelling*, **17**, 31-57.

Thieme, H. R. and Castillo-Chavez, C. (1989). On the role of variable infectivity in the dynamics of the human immunodeficiency virus epidemic. In: *Mathematical and Statistical Approaches to AIDS Epidemiology*. C. Castillo-Chavez and S. A. Levin (Eds.). Lecture Notes in Biomathematics, **83**, 157-176. Berlin; Springer-Verlag.

UNAIDS (1998). AIDS epidemic update: December 1998. *UNAIDS Joint United Nations Programme on HIV/AIDS.*

UNAIDS (1999). Drug Use and HIV/AIDS. UNAIDS statement presented at the United Nations General Assembly special session on drugs. *UNAIDS Joint United Nations Programme on HIV/AIDS.*

Varga, R. S. (1962). *Matrix Iterative Analysis.* Englewood Cliffs, NJ; Prentice Hall Press.

Ward, J. W., Deppe, D. A. and Sason, S. (1987). Risk of HIV infection from blood donors who later developed AIDS. *Annals of Internal Medicine*, **106**, 61-62.

Weiss, R. A. (1985). Human T-cell retroviruses. In: *RNA tumour viruses*, (Ed. R. A. Weiss, N. M. Teich, H. E. Varmus and J. Coffin), 406-485. New York; Cold Spring Harbor Laboratory.

Whittle, P. (1955). The outcome of a stochastic epidemic-a note on Bailey's paper. *Biometrika*, **42**, 116-122.

Yakowitz. S. J. (1994). From a microcosmic IDU model to a macrocosmic HIV epidemic. In: *Modeling the AIDS Epidemic: Planning, Policy and Prediction*, Kaplan, E. H. and Brandeau, M. L. (Eds.), 365-386. New York; Raven Press.

Appendix A

Proof of Theorem 2.5

Simplifying $\det(\mathbf{J} - \omega\mathbf{I}) = 0$ and collecting terms in ω^3 gives

$$\begin{aligned} a_1 &= \mu + \delta_1 + \lambda\alpha\beta^\star(1-\phi) + \lambda\gamma\left[\hat{\tau} + \hat{\theta} + \pi^\star(1-\hat{\theta})\right] + \mu + \delta_2 + \mu + \delta_3, \\ &= \mu + \delta_1 + \mu + \delta_2 + \mu + \delta_3 + \lambda\gamma\frac{\pi^\star}{\beta^\star} + \lambda\alpha\beta^\star(1-\phi), \end{aligned} \tag{A.1}$$

since from eqn (2.7) we can replace $\beta^\star(\hat{\tau} + \hat{\theta} + \pi^\star(1-\hat{\theta}))$ with $\pi^\star$. Similarly collecting terms in ω^2 gives

$$\begin{aligned} a_2 &= \lambda\alpha(1-\phi)\beta^\star\left[\delta_1 + \mu + \delta_2 + \mu + \delta_3\right] \\ &\quad + \lambda\gamma\left[\hat{\tau} + \hat{\theta} + \pi^\star(1-\hat{\theta})\right]\left(\mu + \delta_1 + \mu + \delta_2 + \mu + \delta_3\right) \\ &\quad + (\mu+\delta_1)(\mu+\delta_2) + (\mu+\delta_2)(\mu+\delta_3) + (\mu+\delta_3)(\mu+\delta_1) \\ &\quad + \lambda\alpha(1-\phi)\beta^\star\lambda\gamma\left[\hat{\tau} + \hat{\theta} + \pi^\star(1-\hat{\theta})\right] \\ &\quad - \left(1 - \beta^\star(1-\hat{\theta})\right)\lambda\gamma(1-\pi^\star)\lambda\alpha(1-\phi). \end{aligned}$$

Again from eqn (2.7) we can replace $(\hat{\tau} + \hat{\theta} + \pi^\star(1-\hat{\theta}))$ with $\pi^\star/\beta^\star$ and similarly we can replace $[1 - \beta^\star(1-\hat{\theta})](1-\pi^\star)$ with $1 - \beta^\star(1+\hat{\tau})$. From eqns (2.6) and (2.7) we can also replace $(\mu+\delta_1)(\hat{\tau} + \hat{\theta} + \pi^\star(1-\hat{\theta}))$ with $L(1-\pi^\star)\lambda\alpha(1-\phi)$. Replacing these various terms and after some simplification we find that

$$a_2 = \lambda\alpha(1-\phi)\beta^\star[\delta_1 + \mu + \delta_2 + \mu + \delta_3] \tag{A.2}$$

$$+\lambda\gamma\frac{\pi^\star}{\beta^\star}[\mu + \delta_2 + \mu + \delta_3] \tag{A.3}$$

$$+\lambda\gamma\lambda\alpha(1-\phi)\beta^\star(1+\hat{\tau}) \tag{A.4}$$

$$+(\mu+\delta_1)(\mu+\delta_2) + (\mu+\delta_2)(\mu+\delta_3) + (\mu+\delta_3)(\mu+\delta_1) \tag{A.5}$$

$$+\left[\frac{\delta_1}{\mu+\delta_2} + \frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right](1-\pi^\star)\lambda\alpha(1-\phi)\lambda\gamma. \tag{A.6}$$

Collecting terms in ω gives

$$\begin{aligned} a_3 &= -\Big[1-\beta^\star(1-\hat\theta)\Big]\lambda\gamma(1-\pi^\star)\lambda\alpha(1-\phi)\Big[\delta_1+\mu+\delta_2+\mu+\delta_3\Big] \\ &\quad +\lambda\beta^\star\alpha(1-\phi)L(\mu+\delta_2)(\mu+\delta_3)+(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3) \\ &\quad +(\mu+\delta_2)(\mu+\delta_3)\lambda\gamma\Big(\hat\tau+\hat\theta+\pi^\star(1-\hat\theta)\Big)+\delta_1\lambda\beta^\star\alpha(1-\phi)\lambda\gamma\Big(\hat\tau+\hat\theta+\pi^\star(1-\hat\theta)\Big) \\ &\quad +\Big[\mu+\delta_1+\lambda\alpha\beta^\star(1-\phi)\Big]\lambda\gamma\Big(\hat\tau+\hat\theta+\pi^\star(1-\hat\theta)\Big)\Big[\mu+\delta_2+\mu+\delta_3\Big]. \end{aligned}$$

In a similar manner to the a_2 term above we find that

$$\begin{aligned} a_3 &= \lambda\beta^\star\alpha(1-\phi)L(\mu+\delta_2)(\mu+\delta_3)+(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3) \\ &\quad +(\mu+\delta_2)(\mu+\delta_3)\lambda\gamma\frac{\pi^\star}{\beta^\star}+\lambda\gamma\lambda\alpha(1-\phi)[\delta_1+\mu+\delta_2+\mu+\delta_3]\beta^\star(1+\hat\tau) \\ &\quad +(1-\pi^\star)\lambda\alpha(1-\phi)\lambda\gamma\left[\frac{\delta_1\delta_2}{\mu+\delta_3}+\frac{\delta_1\delta_2}{\mu+\delta_2}+\frac{\delta_1(\mu+\delta_3)}{\mu+\delta_2}\right], \end{aligned} \tag{A.7}$$

and finally the constant term in the characteristic equation is

$$\begin{aligned} a_4 &= -\Big[1-\beta^\star(1-\hat\theta)\Big]\lambda\gamma(1-\pi^\star)\lambda\alpha(1-\phi)L(\mu+\delta_2)(\mu+\delta_3) \\ &\quad +\lambda\alpha\beta^\star(1-\phi)\lambda\gamma\Big[\hat\tau+\hat\theta+\pi^\star(1-\hat\theta)\Big]L(\mu+\delta_2)(\mu+\delta_3) \\ &\quad +(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)\lambda\gamma\Big[\hat\tau+\hat\theta+\pi^\star(1-\hat\theta)\Big]. \end{aligned} \tag{A.8}$$

Again making substitutions similar to those in the previous a_i terms we find that

$$a_4 = \lambda\gamma\lambda\alpha(1-\phi)L(\mu+\delta_2)(\mu+\delta_3)\beta^\star(1+\hat\tau). \tag{A.9}$$

From eqn (2.10) we have that $\beta^\star(1+\hat\tau)=1-(1/R_0)$, hence $a_4>0$ if $R_0>1$. It is clear that for a_1, a_2 and a_3 to be strictly positive it is sufficient that $(\pi^\star,\beta^\star)$ is strictly positive which is true when $R_0>1$.

We now consider $a_1a_2-a_3$, this can be written as

$$\begin{aligned} &\lambda\alpha\beta^\star(1-\phi)\Big[(A.2)+(A.3)+(A.4)+(A.6)+\mu(\mu+\delta_3)+\mu(\mu+\delta_1+\delta_2)\Big] \\ &+(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)\Big[(A.2)+(A.3)+(\mu+\delta_1)(\mu+\delta_2)+(\mu+\delta_1)(\mu+\delta_3)\Big] \\ &+\mu(A.4)+(\mu+\delta_2)(\mu+\delta_3)(\mu+\delta_2+\mu+\delta_3) \\ &+\left[(\mu+\delta_1)\left[\frac{\delta_1}{\mu+\delta_2}+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}\right]+\delta_1\right](1-\pi^\star)\lambda\alpha(1-\phi)\lambda\gamma \\ &+\lambda\gamma\frac{\pi^\star}{\beta^\star}\Big[(A.2)+(A.3)+(A.4)+(A.6)+(\mu+\delta_1)(\mu+\delta_2)+(\mu+\delta_1)(\mu+\delta_3)\Big]. \end{aligned}$$

Therefore $a_1a_2 - a_3 > 0$. We require to show that $(a_1a_2 - a_3)a_3 > a_1^2a_4$. Since $\beta^\star(1+\hat{\tau})$ is a factor of $a_1^2a_4$ we shall base our argument around showing that $(a_1a_2 - a_3)a_3$ has sufficient terms containing $\beta^\star(1+\hat{\tau})$ such that $(a_1a_2 - a_3)a_3 > a_1^2a_4$. It is sufficient to show that

$$\left\{\left[\delta_1(1-\pi^\star)\lambda\alpha(1-\phi)\lambda\gamma + (A.2)\lambda\gamma\frac{\pi^\star}{\beta^\star} + (\mu+\delta_1)(\mu+\delta_2+\mu+\delta_3)\lambda\gamma\frac{\pi^\star}{\beta^\star}\right]\right.$$

$$\left.\times(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)\right\}$$

$$+\left\{\left[\lambda\alpha\beta^\star(1-\phi) + \mu + \lambda\gamma\frac{\pi^\star}{\beta^\star}\right]\beta^\star(1+\hat{\tau})\lambda\alpha(1-\phi)\lambda\gamma\right.$$

$$\left.\times\Big(a_3 - \lambda\gamma\lambda\alpha(1-\phi)[\delta_1+\mu+\delta_2+\mu+\delta_3]\beta^\star(1+\hat{\tau})\Big)\right\}$$

$$+(a_1a_2 - a_3)(\delta_1+\mu+\delta_2+\mu+\delta_3)\beta^\star(1+\hat{\tau})\lambda\alpha(1-\phi)\lambda\gamma$$

$$> \quad a_1^2\Big[(\mu+\delta_2)(\mu+\delta_3) + \delta_1(\mu+\delta_3) + \delta_1\delta_2\Big]\beta^\star(1+\hat{\tau})\lambda\alpha(1-\phi)\lambda\gamma. \qquad (A.10)$$

Consider the term in the first square bracket in inequality (A.10)

$$\delta_1(1-\pi^\star)\lambda\alpha(1-\phi)\lambda\gamma + \lambda\gamma\lambda\alpha(1-\phi)(\delta_1+\mu+\delta_2+\mu+\delta_3)\pi^\star$$

$$+(\mu+\delta_2+\mu+\delta_3)\lambda\gamma(\mu+\delta_1)\frac{\pi^\star}{\beta^\star}$$

$$= \lambda\alpha(1-\phi)\lambda\gamma\Big[\delta_1(1-\pi^\star) + (\delta_1+\mu+\delta_2+\mu+\delta_3)\pi^\star + (\mu+\delta_2+\mu+\delta_3)L(1-\pi^\star)\Big]$$

$$> \lambda\alpha(1-\phi)\lambda\gamma(\delta_1+\mu+\delta_2+\mu+\delta_3)\beta^\star(1+\hat{\tau}), \qquad \text{as } \beta^\star(1+\hat{\tau}) < 1. \qquad (A.11)$$

Hence using inequality (A.11) it is sufficient to show that

$$(\delta_1+\mu+\delta_2+\mu+\delta_3)(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)$$

$$+\left(\lambda\alpha\beta^\star(1-\phi) + \mu + \lambda\gamma\frac{\pi^\star}{\beta^\star}\right)\Big(a_3 - \lambda\gamma\lambda\alpha(1-\phi)[\delta_1+\mu+\delta_2+\mu+\delta_3]\beta^\star(1+\hat{\tau})\Big)$$

$$+(a_1a_2 - a_3)(\delta_1+\mu+\delta_2+\mu+\delta_3) \qquad (A.12)$$

$$> \Big\{(\mu+\delta_2)(\mu+\delta_3) + \delta_1(\mu+\delta_3) + \delta_1\delta_2\Big\} \times$$

$$\Big\{(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)^2 \qquad (A.13)$$

$$+\left(\frac{\lambda\gamma\pi^\star}{\beta^\star}\right)^2 \qquad (A.14)$$

$$+[\lambda\alpha\beta^\star(1-\phi)]^2 \qquad (A.15)$$

$$+2(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)\frac{\lambda\gamma\pi^\star}{\beta^\star} \tag{A.16}$$

$$+2(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)\lambda\alpha\beta^\star(1-\phi) \tag{A.17}$$

$$+2\lambda\gamma\pi^\star\lambda\alpha(1-\phi)\Big\}. \tag{A.18}$$

It is straightforward to show that

$$(\delta_1+\mu+\delta_2+\mu+\delta_3)(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)+\mu(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)$$

$$+(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)(\mu+\delta_1)(\mu+\delta_2+\mu+\delta_3)(\delta_1+\mu+\delta_2+\mu+\delta_3)$$

$$+(\mu+\delta_2)(\mu+\delta_3)(\mu+\delta_2+\mu+\delta_3)(\delta_1+\mu+\delta_2+\mu+\delta_3)$$

$$> \quad (\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)^2\Big[(\mu+\delta_2)(\mu+\delta_3)+\delta_1(\mu+\delta_3)+\delta_1\delta_2\Big]. \tag{A.19}$$

Hence the term in (A.13) can be cancelled by the terms in (A.12) containing only μ, δ_1, δ_2 and δ_3. Similarly

$$(\mu+\delta_2)(\mu+\delta_3)+(\mu+\delta_2+\mu+\delta_3)(\delta_1+\mu+\delta_2+\mu+\delta_3)$$

$$> \quad \Big[(\mu+\delta_2)(\mu+\delta_3)+\delta_1(\mu+\delta_3)+\delta_1\delta_2\Big], \tag{A.20}$$

and (A.14) can be cancelled by the terms in (A.12) containing $(\lambda\gamma\pi^\star/\beta^\star)^2$. The term in (A.15) can be found explicitly in (A.12). Moreover (A.16) can be cancelled by the terms in (A.12) containing $\lambda\gamma\pi^\star/\beta^\star$ as

$$(\mu+\delta_1)(\mu+\delta_2)(\mu+\delta_3)+\mu(\mu+\delta_2)(\mu+\delta_3)$$

$$+(\delta_1+\mu+\delta_2+\mu+\delta_3)(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)(\mu+\delta_2+\mu+\delta_3)$$

$$+(\delta_1+\mu+\delta_2+\mu+\delta_3)(\mu+\delta_1)(\mu+\delta_2+\mu+\delta_3)$$

$$> \quad 2(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)\Big[(\mu+\delta_2)(\mu+\delta_3)+\delta_1(\mu+\delta_2+\delta_3)\Big]. \tag{A.21}$$

In a similar fashion

$$(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)(\delta_1+\mu+\delta_2+\mu+\delta_3)^2$$

$$> \quad 2(\mu+\delta_1+\mu+\delta_2+\mu+\delta_3)\Big[(\mu+\delta_2)(\mu+\delta_3)+\delta_1(\mu+\delta_3)+\delta_1\delta_2\Big]. \tag{A.22}$$

Hence (A.17) can be cancelled by the terms in (A.12) containing $\lambda\alpha(1-\phi)\beta^\star$. Finally the terms in (A.12) containing $\lambda\gamma\pi^\star\lambda\alpha(1-\phi)$ will cancel (A.18) as

$$(\mu+\delta_2)(\mu+\delta_3)L+(\mu+\delta_2+\mu+\delta_3)(\delta_1+\mu+\delta_2+\mu+\delta_3)+(\delta_1+\mu+\delta_2+\mu+\delta_3)^2$$

$$> \quad 2\Big[(\mu+\delta_2)(\mu+\delta_3)+\delta_1(\mu+\delta_3)+\delta_1\delta_2\Big]. \tag{A.23}$$

Hence $a_1a_2a_3 > a_3^2 + a_1^2a_4$ and all the Routh-Hurwitz conditions are satisfied for $R_0 > 1$.•

Appendix B

Parameter Estimation

We do not possess our own source of data from which to estimate the parameters in our various models. We instead rely on estimates from existing published work and adapt these for our own purposes. In their paper, Kaplan and O'Keefe estimate values for all the parameters in their model, however we do not use all these original estimates as later work by various authors (particularly Kaplan) provides some improved estimates.

It is fair to say that there does not appear to be a generic set of parameter estimates available relating to the behaviour of intravenous drug addicts. There have been many studies covering the lifestyle and addiction habits of drug users, (for example Barnard and Frischer, 1995), unfortunately different studies produce markedly different parameter estimates. To compound this problem it is very difficult to assess the reliability of data from drug addicts. Much of the data available comes from self assessment questionnaires which the addicts complete while receiving counselling from outreach workers or health clinics.

The problem of subjective parameter estimates was discussed in Kaplan (1995) who proposed a solution to this problem by using a much simplified model of the infection process. The proposed model was very simple but had the advantage that all the parameters contained in this model could be estimated using objective data collected by a needle exchange scheme, such as the average duration a needle remains in circulation and the fraction of needles which are HIV positive when exchanged. The parameters in this model can be thought of as aggregate forms of some of the parameters contained in our models. By breaking these aggregate forms down into their constituent parts we arrive at some of the estimates that we need.

In early models of the spread of HIV among intravenous drug users, (Kaplan, 1989a), the probability that an infectious needle was flushed by an uninfectious addict was featured explicitly. However in later work (Kaplan and Heimer, 1992a) it was assumed

that this event could never occur. We follow the latter work and take $\theta = 0$. Whilst it is not really known whether flushing occurs assuming that it does not will ensure that our models do not underestimate the prevalence of HIV infection. Also authors other than Kaplan do not include flushing in their models (Peterson et al., 1990, Blower et al., 1991, Kretzschmar and Wiessing, 1998). Further supporting evidence for the assumption that $\theta = 0$ is given by Kaplan (private communication, 1996) who states that HIV can be detected in syringes used by infectious addicts in which the blood remaining in the syringe after the addict has used it has been very greatly diluted.

Estimating the fraction of addicts who successfully clean needles prior to use is difficult. The difficulty arises not due to lack of data but to the very varied data available, for example at the lower end of the estimates we have Goldberg et al. (1995) who estimate that $\phi = 0.44$ from a survey of drug users in Glasgow, UK. At the higher end of the estimates we have Kaplan and O'Keefe's estimate of $\phi = 0.84$. This estimate was calculated from questionnaire data from attenders at a needle exchange. This is an environment where addicts are subject to messages about cleaning their needles before use so it is not clear how reliable the estimate is. Kaplan and O'Keefe express concern that this estimate may be too high and their sensitivity analyses suggest ϕ in the range of 0.42-0.84. In the absence of a more rigorous approach we adopt the middle ground and take $\phi = 0.64$.

Kaplan and O'Keefe (1993), Kaplan (1994) and related papers contain data derived from the New Haven needle exchange program. In Kaplan and O'Keefe (1993) it was assumed that in the absence of a formal exchange program needles circulated indefinitely. We wish to estimate the natural needle turnover rate, that is the rate at which needles are exchanged in the absence of a formal needle exchange program. It is unrealistic to assume that needles remain in circulation for all time since at the very least they have a limited working lifetime. Using data collected from the New Haven needle exchange program Kaplan (1995) estimates that the natural needle turnover rate is $\tau = 15.53$ per year, so the working lifetime of a needle is 23.50 days.

The homogeneous shared needle injection rate is one of the main deficiencies of the basic Kaplan and O'Keefe model. For simplicity it was assumed that all addicts inject at the same rate, this assumption is contrary to observed evidence which suggests that addicts inject at a wide variety of different rates, (see Greenhalgh (1996, 1997) for models which incorporate heterogeneity in the needle sharing rate). Kaplan and O'Keefe (1993) estimate the shared injection rate for New Haven addicts to be $\lambda =$

246.22 per year. Data from Goldberg et al. (1995) suggest a mean shared injection rate of only $\lambda = 72.48$ per year for drug addicts in Glasgow. However Goldberg et al. report that the distribution of needle sharing is highly skew in that many addicts do not share at all, whilst a small minority share equipment between 900-1800 times a year. Also Goldberg et al. obtain their information on all drug users in Glasgow whilst Kaplan and O'Keefe consider attenders at a needle exchange scheme and the latter group may well share needles more frequently. Given this wide range of sharing rates it seems that Kaplan and O'Keefe's estimate is not unreasonable. Intuitively it makes sense to overestimate rather than underestimate this parameter since it easy to argue that a small minority of very high risk users will have a disproportionally large effect on the spread of the disease. This intuition is confirmed by analytical results (Greenhalgh, 1996, 1997). We again follow Kaplan and O'Keefe and take $\lambda = 246.22$ per year.

Kaplan (1995) uses a model which deals only with infectious needles and where the parameters can be estimated entirely from objective data. Kaplan estimates that the rate at which uncontaminated needles become contaminated, $\tilde{\lambda}$, is 0.3675 per day, and the rate at which contaminated needles become uncontaminated, $\tilde{\mu}$, is 0.1665 per day. Note that these estimates assume that needles are never exchanged (a similar assumption was made in Kaplan and O'Keefe, 1993). In our notation we have that

$$\tilde{\lambda} = (\lambda/365)\gamma\pi^{\star} = 0.3675, \tag{B.1}$$

and

$$\tilde{\mu} = (\lambda/365)\gamma(1-\pi^{\star})\phi = 0.1665. \tag{B.2}$$

Eqn (B.1) corresponds to the event where an infected addict injects with a randomly chosen needle and leaves it contaminated (where the population is in equilibrium). Eqn (B.2) corresponds to the event where a contaminated needle is used by an uninfected addict and is cleaned prior to use, rendering the needle uncontaminated. We have assumed that θ, the probability that an uninfected addict flushes a needle, is zero as previously estimated. Equating eqns (B.1)-(B.2) gives us

$$\frac{\pi^{\star}}{(1-\pi^{\star})\phi} = 2.21. \tag{B.3}$$

Kaplan (1995) estimates that the prevalence of HIV in the population prior to any external intervention (such as needle exchange) is approximately 0.6. We have already estimated that addicts successfully clean needles prior to use with probability 0.64. Substituting these values into the left hand side of eqn (B.3) gives us a value of 2.34

which is close to the observed value of 2.21 which is a reasonable assurance that our individual estimates of $\pi^\star = 0.6$ and $\phi = 0.64$ are realistic. Now consider eqn (B.1), using our estimate of $\pi^\star$ we find that $\lambda\gamma = 223.5625$, we estimate that $\lambda = 246.22$ which implies that the gallery ratio, γ, is approximately 0.908. This means that we have roughly 908 addicts for every 1000 needles. Note that this estimate is five times greater than the estimate of 0.1675 given in Kaplan and O'Keefe (1993). However we believe that our estimate is probably more accurate than the Kaplan and O'Keefe value as the latter estimate is based on needle exchange data collected over a relatively short period from November 1990-February 1991, whereas the estimates in Kaplan (1995) are based on data over a considerably longer period from November 1990-June 1992.

Longini et al. (1991) estimate that the average duration of the AIDS incubation period is 117.7 months (9.808 years). Hence we take $\delta = 0.1020$ per year. They also estimate that the average durations an addict spends in state one (Acute Infection), state two (Asymptomatic) and state three (Pre-AIDS Symptoms) are 2.6, 52.6 and 62.5 months respectively. Hence the average durations in each stage in years are 0.2166, 4.3833 and 5.2083 respectively, taking the reciprocal of these values gives us $\delta_1 = 4.6154$ per year, $\delta_2 = 0.2281$ per year and $\delta_3 = 0.192$ per year. In our models the parameter μ represents the rate at which addicts leave the needle sharing population for reasons other than developing full blown AIDS. We follow Caulkins and Kaplan (1991) who estimate that $\mu = 0.1333$ per year.

We also wish to estimate the probability of being infected by HIV after injecting once with a contaminated needle. The has already been dealt with in the literature, again by Kaplan and O'Keefe (1993) who estimate $\alpha = 0.0066$. The particular difficulty with estimating this parameter is the lack of data, the only data available relating to this subject is the level of "needle-stick injuries". These are accidents involving the infection of health workers from contaminated needles in their possession. The fraction of these accidents which resulted in the worker becoming infected with HIV was in the region of 0.003-0.005, (Kaplan and O'Keefe, 1993). This estimate can only serve as a lower bound on the value of α since the risk of transmission is certainly much greater to addicts who inject with a contaminated needle rather than merely pricking themselves by accident, as the volume of blood transferred in a needlestick injury is typically much smaller than that transferred when an addict injects with a previously used syringe.

To overcome this problem Kaplan and O'Keefe used a model based estimation technique to compute α. They estimated all model parameters with the exception

of α, they also estimated the prevalence of contaminated needles in circulation, this value could be estimated accurately using needle exchange scheme data. Under the assumption that the population is in a steady state and that the model is a good representation of reality then the endemic equilibrium level of needles should be equal to the observed value. In other words Kaplan and O'Keefe set the expression for the endemic level of needles in their model equal to the observed value and solved to find α. An obvious problem with this method is that a different model will produce a different estimate of α, which is not ideal, however in the absence of a better alternative we adopt the same method as Kaplan and O'Keefe to estimate α.

Consider the Simple Model, from the equilibrium version of eqn (2.4) we have that

$$\pi^\star = \frac{\beta^\star(\hat{\theta} + \hat{\tau})}{1 - \beta^\star(1 - \hat{\theta})}. \tag{B.4}$$

From eqn (2.6) we find that

$$\alpha = \frac{\pi^\star(\mu + \delta_1)}{(1 - \pi^\star)\lambda\beta^\star(1 - \phi)L}, \tag{B.5}$$

hence we have that

$$\alpha = \frac{(\hat{\theta} + \hat{\tau})(\mu + \delta_1)}{[1 - \beta^\star(1 + \hat{\tau})]\lambda(1 - \phi)L}. \tag{B.6}$$

We know all parameters on the right hand side of equation (B.6) and we take $\beta^\star = 0.675$ as estimated by Kaplan and O'Keefe. Using the estimates outlined previously we arrive at a value of $\alpha = 0.00601$.

Using data on viral antigen recovery and epidemiological data from transfusion recipients Peterson et al. (1990) estimate that the relative viral load of addicts exists in the ratio 5:1:3 for Acute Infection:Asymptomatic:Pre-AIDS Symptoms. There has been some recent work (Koopman et al., 1997, Hyman and Stanley, 1999) which suggests more extreme ratios, however we use Peterson's estimates as they seem well founded on medical literature. Hence we assume that a needle used by an addict in stage one (Acute Infection) will be left fives times more infectious than if the addict were in stage two (Asymptomatic). Similarly we assume that a needle used by an addict in stage three (Pre-AIDS Symptoms) will be three times more infectious than if the addict were in stage two. We therefore assume that $\alpha_1 = \zeta_1\alpha_2$, $\alpha_3 = \zeta_3\alpha_2$ and that α_1:α_2:α_3 exist in the ratio 5:1:3. Hence we only need to estimate the baseline transmission probability, α_2. We use the same method as that used to estimate α in the Simple Model in Chapter 2. We estimate all other model parameters and the endemic fraction

of infected needles and solve for α_2 in the expression for $\beta^\star$. An additional complication compared to the Simple Model case is that we are using three models, the Optimistic Model, the Pessimistic Model and the General Model to examine the effect of three stage infectivity. The method of estimation depends on the model being used, therefore using the Optimistic Model to estimate α_2 will give us a different estimate from using the Pessimistic Model or the General Model to estimate α_2. This is obviously not sensible as α_2 has the same physical interpretation in all three of our models and therefore we should use the same estimate in simulations of each model.

It was decided to use the Pessimistic Model rather than the Optimistic Model or General Model to estimate α_2. The reason for this is that we feel it is more realistic to assume that a susceptible addict never flushes a needle than always flushing a needle, which suggests that the Pessimistic Model may be more realistic than the Optimistic Model. However in truth neither the Optimistic or Pessimistic Model is ideal for the purpose of estimating α_2 since by construction both models will give a biased estimate of the true value. Ideally we should use the General Model (which is discussed in Chapter 5), however this is not straightforward as this model is much more complex than the Optimistic and Pessimistic Models and contains a number of parameters which we cannot estimate with any kind of accuracy. Hence we settle for the Pessimistic Model as our most practical method of estimating a realistic value for α_2. Proceeding with the Pessimistic Model we find that

$$\pi^\star = \frac{\beta^\star(\phi+\hat{\tau})}{1-\beta^\star(1-\phi)}, \tag{B.7}$$

and

$$\beta_1^\star = \frac{\pi^\star}{(1-\phi)\pi^\star + L(\phi+\hat{\tau})}, \tag{B.8}$$

using eqns (4.4)-(4.6) and replacing $\pi_1^\star$ with $\pi^\star/L$ in eqn (4.4). Using eqn (B.7) to replace $\pi^\star$ in eqn (B.8) gives us

$$\beta_1^\star = \frac{\beta^\star}{\beta^\star(1-\phi) + L[1-\beta^\star(1-\phi)]}. \tag{B.9}$$

Adding eqns (4.4) and (4.6) we get

$$\beta_{1+3}^\star = \frac{\lambda\gamma\pi_{1+3}^\star}{(1-\phi)\lambda\gamma\pi_{1+3}^\star + (\phi\lambda\gamma+\tau)}, \tag{B.10}$$

$$= \frac{\beta^\star}{\beta^\star(1-\phi) + \frac{L(1-\beta^\star(1-\phi))}{1+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}}}, \tag{B.11}$$

by using eqn (B.7) to replace $\pi^\star$. From the equilibrium version of eqn (4.1) and again

substituting in the expression for $\pi^\star$ from eqn (B.7) we have that

$$\alpha_2 = \frac{\beta^\star(\phi+\hat{\tau})(\mu+\delta_1)}{[1-\beta^\star(1+\hat{\tau})]\lambda(1-\phi)(\beta_1^\star(\zeta_1-\zeta_3)+\beta^\star+\beta_{1+3}^\star(\zeta_3-1))L}, \quad \text{(B.12)}$$

where $\alpha_1 = \zeta_1\alpha_2$ and $\alpha_3 = \zeta_3\alpha_2$. Substituting in the previous expressions for $\beta_1^\star$ and $\beta_{1+3}^\star$ gives us

$$\alpha_2 = \frac{(\phi+\hat{\tau})(\mu+\delta_1)}{L[1-\beta^\star(1+\hat{\tau})]\lambda(1-\phi)\{1+X+Y\}}, \quad \text{(B.13)}$$

where

$$X = \frac{\zeta_1-\zeta_3}{\beta^\star(1-\phi)+L[1-\beta^\star(1-\phi)]}, \quad \text{(B.14)}$$

and

$$Y = \frac{\zeta_3-1}{\beta^\star(1-\phi)+\frac{L[1-\beta^\star(1-\phi)]}{1+\frac{\delta_1\delta_2}{(\mu+\delta_2)(\mu+\delta_3)}}}. \quad \text{(B.15)}$$

The expression in eqn (B.13) is complicated but using the parameter estimates from Table 3.1 we know the values of all the component parts and $\beta^\star = 0.675$, hence we have an estimate of $\alpha_2 = 0.002824$. Using the estimates of $\zeta_1 = 5$ and $\zeta_3 = 3$ we have that $\alpha_1 = 0.014121$ and $\alpha_3 = 0.008473$.

Appendix C

Endemic Stability in the Optimistic Model

We briefly demonstrate that if $\lambda\gamma$ is large compared with the other parameters of the model apart from τ, (including $\lambda\alpha_1$, $\lambda\alpha_2$ and $\lambda\alpha_3$) then the endemic equilibrium in the Optimistic Model is locally stable. Firstly the characteristic equation of the Jacobian matrix in the full Optimistic Model about the endemic equilibrium is $\det \mathbf{A} = 0$ where

$$\mathbf{A} = \begin{bmatrix}
-(\mu+\delta_1)-\lambda\hat{\beta}^*(1-\phi)-\omega & -\lambda\hat{\beta}^*(1-\phi) & -\lambda\hat{\beta}^*(1-\phi) & (1-\pi^*)\lambda\alpha_1(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\
\delta_1 & -(\mu+\delta_2)-\omega & 0 & 0 & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3)-\omega & 0 & 0 & 0 \\
\lambda\gamma & 0 & 0 & -(\lambda\gamma+\tau)-\omega & 0 & 0 \\
0 & \lambda\gamma & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda\gamma & 0 & -(\lambda\gamma+\tau)-\omega & 0
\end{bmatrix},$$

and $\hat{\beta}^\star = \beta_1^\star\alpha_1 + \beta_2^\star\alpha_2 + \beta_3^\star\alpha_3$. Now since this determinant is unaffected by row and column operations we have that $\det \mathbf{A} = \det \mathbf{A}_1$ where $\mathbf{A}_1$ is $\mathbf{A}$ with column i now replaced by column i + {column $(i+3)$}$/(1+\hat{\tau})$ for $i = 1, 2, 3$. Columns 4, 5 and 6 in $\mathbf{A}_1$ are the same as in $\mathbf{A}$.

Hence $\mathbf{A}_1 =$

$$\begin{bmatrix}
-(\mu+\delta_1)-\lambda\hat{\beta}^*(1-\phi)+\frac{\lambda\alpha_1(1-\phi)(1-\pi^*)}{1+\hat{\tau}}-\omega & -\lambda\hat{\beta}^*(1-\phi)+\frac{\lambda\alpha_2(1-\phi)(1-\pi^*)}{1+\hat{\tau}} & -\lambda\hat{\beta}^*(1-\phi)+\frac{\lambda\alpha_3(1-\phi)(1-\pi^*)}{1+\hat{\tau}} \\
\delta_1 & -(\mu+\delta_2)-\omega & 0 \\
0 & \delta_2 & -(\mu+\delta_3)-\omega \\
-\frac{\omega}{1+\hat{\tau}} & 0 & 0 \\
0 & -\frac{\omega}{1+\hat{\tau}} & 0 \\
0 & 0 & -\frac{\omega}{1+\hat{\tau}}
\end{bmatrix}$$

$$\begin{bmatrix}
(1-\pi^*)\lambda\alpha_1(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\
0 & 0 & 0 \\
0 & 0 & 0 \\
-(\lambda\gamma+\tau)-\omega & 0 & 0 \\
0 & -(\lambda\gamma+\tau)-\omega & 0 \\
0 & 0 & -(\lambda\gamma+\tau)-\omega
\end{bmatrix}.$$

We have that $\det \mathbf{A}_1 = (\lambda\gamma+\tau+\omega)^3 \det \mathbf{B}_1$ plus terms involving at most two factors $(\lambda\gamma+\tau+\omega)^2$, where $\det \mathbf{B}_1 = 0$ represents the characteristic equation of the "addict only" Optimistic Model about its endemic equilibrium. So if $\lambda\gamma \to \infty$ with $\hat{\tau} = \tau/(\lambda\gamma)$ fixed then

$$\frac{\det \mathbf{A}_1}{(\lambda\gamma+\tau+\omega)^3 \det \mathbf{B}_1} \to 1.$$

Therefore the roots of the characteristic equation of the full model about its endemic equilibrium tend to $-(\lambda\gamma+\tau)$, $-(\lambda\gamma+\tau)$,$-(\lambda\gamma+\tau)$ and the roots of the characteristic equation of the "addict only" model. Hence if $\lambda\gamma$ is large enough then all eigenvalues have strictly negative real parts and the endemic equilibrium in the full model is locally stable when $R_0 > 1$.

Appendix D

Endemic Stability in the Pessimistic Model

D.1 Local Stability

We first show that as in the Optimistic Model if $\lambda\gamma$ is large compared with the other parameters of the model apart from τ, (including $\lambda\alpha_1$, $\lambda\alpha_2$ and $\lambda\alpha_3$) then the endemic equilibrium in the Pessimistic Model is locally stable. Using the coordinate system $(\pi_1, \pi_2, \pi_3, \beta_1, \beta, \beta_{1+3})$ the characteristic equation of the Jacobian matrix in the full Pessimistic Model about the endemic equilibrium is $\det \mathbf{A} = 0$ where $\mathbf{A} =$

$$\begin{bmatrix}
-\lambda(1-\phi)\beta^\dagger-(\mu+\delta_1)-\omega & -\lambda(1-\phi)\beta^\dagger & -\lambda(1-\phi)\beta^\dagger & \lambda(1-\phi)(\alpha_1-\alpha_3)(1-\pi^*) \\
\delta_1 & -(\mu+\delta_2)-\omega & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3)-\omega & 0 \\
\lambda\gamma(1-\beta_1^*(1-\phi)) & 0 & 0 & -\lambda\gamma(\pi_1^*(1-\phi)+\phi+\hat{\tau})-\omega \\
0 & \lambda\gamma(1-\beta^*(1-\phi)) & 0 & 0 \\
0 & 0 & \lambda\gamma(1-\beta_{1+3}^*(1-\phi)) & 0
\end{bmatrix}$$

$$\begin{bmatrix}
\lambda\alpha_2(1-\phi)(1-\pi^*) & \lambda(\alpha_3-\alpha_2)(1-\phi)(1-\pi^*) \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
-\lambda\gamma(\pi^*(1-\phi)+\phi+\hat{\tau})-\omega & 0 \\
0 & -\lambda\gamma(\pi_{1+3}^*(1-\phi)+\phi+\hat{\tau})-\omega
\end{bmatrix},$$

and $\beta^\dagger = \beta_1^\star(\alpha_1 - \alpha_3) + \beta^\star\alpha_2 + \beta_{1+3}^\star(\alpha_3 - \alpha_2)$. As in the corresponding result for the Optimistic Model we wish to construct a matrix $\mathbf{A}_1$, where $\det(\mathbf{A}) = \det(\mathbf{A}_1)$ and where rows 1-3 and columns 1-3 in $\mathbf{A}_1$ correspond to the Jacobian in the "addict only" Pessimistic Model. First considering the top left term in $\mathbf{A}$, a_{11}, we have that

$$a_{11} = -\lambda(1-\phi)\beta^\dagger - (\mu+\delta_1) - \omega,$$

$$= -\lambda(1-\phi)\left[\frac{\pi_1^\star(\alpha_1-\alpha_3)}{\pi_1^\star(1-\phi)+\hat{\tau}+\phi}+\frac{\pi^\star\alpha_2}{\pi^\star(1-\phi)+\hat{\tau}+\phi}+\frac{\pi_{1+3}^\star(\alpha_3-\alpha_2)}{\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi}\right]$$

$$-(\mu+\delta_1)-\omega,$$

$$= -\lambda(1-\phi)\left\{\frac{\pi_1^\star(\alpha_1-\alpha_3)[\pi_1^\star(1-\phi)+\hat{\tau}+\phi]}{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)^2}+\frac{\pi^\star\alpha_2[\pi^\star(1-\phi)+\hat{\tau}+\phi]}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}\right.$$

$$\left.+\frac{\pi_{1+3}^\star(\alpha_3-\alpha_2)[\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi]}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2}\right\}-(\mu+\delta_1)-\omega.$$

The top left entry in the Jacobian of eqns (4.31)-(4.33) (the "addict only" pessimistic model) at $(\pi_1^\star,\pi_2^\star,\pi_3^\star)$ is j_{11}. Using eqn (4.34) we find that

$$a_{11}+\omega = j_{11}-\frac{\lambda(1-\phi)(\alpha_1-\alpha_3)(1-\pi^\star)(\hat{\tau}+\phi)}{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)^2}-\frac{\lambda(1-\phi)\alpha_2(1-\pi^\star)(\hat{\tau}+\phi)}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}$$

$$-\frac{\lambda(1-\phi)(\alpha_3-\alpha_2)(1-\pi^\star)(\hat{\tau}+\phi)}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2}.$$

Therefore we can construct a matrix $\mathbf{A_1}$ where the top left entry equals $j_{11}-\omega$ by performing the following column operations on $\mathbf{A}$:

$$\text{col1} = \text{col1}+\text{col4}\times\frac{(\hat{\tau}+\phi)}{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)^2}+\text{col5}\times\frac{(\hat{\tau}+\phi)}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}$$

$$+\text{col6}\times\frac{(\hat{\tau}+\phi)}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2}.$$

In a very similar manner we can produce entries in $\mathbf{A_1}$ corresponding to j_{12} and j_{13} in eqns (4.35) and (4.36) respectively by performing the following column operations on $\mathbf{A}$:

$$\text{col2} = \text{col2}+\text{col5}\times\frac{(\hat{\tau}+\phi)}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}$$

and

$$\text{col3} = \text{col3}+\text{col5}\times\frac{(\hat{\tau}+\phi)}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2}+\text{col6}\times\frac{(\hat{\tau}+\phi)}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2}.$$

This gives us $\mathbf{A_1}=$

$$\begin{bmatrix}
j_{11}-\omega & j_{12} & j_{13} & \lambda(1-\phi)(\alpha_1-\alpha_3)(1-\pi^\star)\\
\delta_1 & -(\mu+\delta_2)-\omega & 0 & 0\\
0 & \delta_2 & -(\mu+\delta_3)-\omega & 0\\
-\frac{(\hat{\tau}+\phi)\omega}{(\pi_1^\star(1-\phi)+\hat{\tau}+\phi)^2} & 0 & 0 & -\lambda\gamma(\pi_1^\star(1-\phi)+\phi+\hat{\tau})-\omega\\
0 & -\frac{(\hat{\tau}+\phi)\omega}{(\pi^\star(1-\phi)+\hat{\tau}+\phi)^2} & 0 & 0\\
0 & 0 & -\frac{(\hat{\tau}+\phi)\omega}{(\pi_{1+3}^\star(1-\phi)+\hat{\tau}+\phi)^2} & 0
\end{bmatrix}$$

$$\begin{bmatrix} (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda(\alpha_3-\alpha_2)(1-\phi) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -\lambda\gamma(\pi^*(1-\phi)+\phi+\hat{\tau})-\omega & 0 \\ 0 & -\lambda\gamma(\pi^*_{1+3}(1-\phi)+\phi+\hat{\tau})-\omega \end{bmatrix}.$$

Using a similar argument as in the corresponding result for the Optimistic Model we have that if $\lambda\gamma \to \infty$ with $\hat{\tau} = \tau/(\lambda\gamma)$ fixed then the roots of the characteristic equation of the full Pessimistic Model about its endemic equilibrium tend to $-\lambda\gamma(\pi^*_1(1-\phi)+\hat{\tau}+\phi)$, $-\lambda\gamma(\pi^*(1-\phi)+\hat{\tau}+\phi)$, $-\lambda\gamma(\pi^*_{1+3}(1-\phi)+\hat{\tau}+\phi)$ and the roots of the characteristic equation of the "addict only" model. Hence if $\lambda\gamma$ is large enough then all eigenvalues have strictly negative real parts and the endemic equilibrium in the full model is locally stable when $R_0 > 1$.

D.2 Lyapunov Stability

We now show that by using different lower bounds to construct the matrix $\mathbf{W}^+$ in the Pessimistic Model (as defined on page 134) we can show that it has eigenvalues whose real parts are all strictly negative, which is true if and only if $\mathbf{W}^+$ is Lyapunov stable. Using Theorem 4.4 we can replace we can replace the variables $\hat{\beta}$, π_1, π_{1+3} and π in $\mathbf{W}(\mathbf{x})$ with lower bounds of $\hat{k}\sigma$, $k_1\sigma$, $k_2\sigma$ and $k_3\sigma$ respectively, where σ is a small strictly positive fixed value. Note that for σ sufficiently small we can choose $\hat{k}$, k_1, k_2 and k_3 to be any positive values. We shall denote this alternative form of $\mathbf{W}^+$ as $\mathbf{W}^+_1$, hence we have that for $t \geq T^+_1$, $\mathbf{W}(\mathbf{x}) \leq \mathbf{W}^+_1 = \mathbf{W}(\mathbf{0}) - \sigma\mathbf{E}_1$ where $\mathbf{E}_1 =$

$$\begin{bmatrix} \lambda(1-\phi)\hat{k} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda\gamma(1-\phi)k_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda\gamma(1-\phi)k_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda\gamma(1-\phi)k_3 \end{bmatrix}.$$

Consider the co-ordinate system $\mathbf{x}' = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}, \tilde{\beta}_{1+3})$. We have that

$$\frac{d\tilde{\mathbf{x}}'}{dt} = \mathbf{V}'(\mathbf{x})\tilde{\mathbf{x}}',$$

where $\mathbf{x}'^T = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\beta}_1, \tilde{\beta}, \tilde{\beta}_{1+3})$ and $\mathbf{V}'(\mathbf{x}) =$

$$\begin{bmatrix}
-(\mu+\delta_1)-\lambda\hat{\beta}(1-\phi) & -\lambda\hat{\beta}(1-\phi) & -\lambda\hat{\beta}(1-\phi) & (1-\pi^*)\lambda(\alpha_1-\alpha_3)(1-\phi) \\
\delta_1 & -(\mu+\delta_2) & 0 & 0 \\
0 & \delta_2 & -(\mu+\delta_3) & 0 \\
\lambda\gamma(1-\beta_1^*(1-\phi)) & 0 & 0 & -\lambda\gamma(\phi+\hat{\tau}+\pi_1(1-\phi)) \\
\lambda\gamma(1-\beta^*(1-\phi)) & \lambda\gamma(1-\beta^*(1-\phi)) & \lambda\gamma(1-\beta^*(1-\phi)) & 0 \\
\lambda\gamma(1-\beta_{1+3}^*(1-\phi)) & 0 & \lambda\gamma(1-\beta_{1+3}^*(1-\phi)) & 0
\end{bmatrix}$$

$$\begin{bmatrix}
(1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda(\alpha_3-\alpha_2)(1-\phi) \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
-\lambda\gamma(\phi+\hat{\tau}+\pi(1-\phi)) & 0 \\
0 & -\lambda\gamma(\phi+\hat{\tau}+\pi_{1+3}(1-\phi))
\end{bmatrix}.$$

We have that $\tilde{\mathbf{x}}' = \mathbf{J_1}\tilde{\mathbf{x}}$ where

$$\mathbf{J_1} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1
\end{bmatrix}.$$

It is straightforward to show that $\mathbf{J_1V} = \mathbf{V'J_1}$, hence if $\mathbf{Ve} = \omega\mathbf{e}$, $\mathbf{V'J_1e} = \omega\mathbf{J_1e}$ so any eigenvalue of $\mathbf{V}$ is an eigenvalue of $\mathbf{V}'$. Similarly any eigenvalue of $\mathbf{V}'$ is an eigenvalue of $\mathbf{V}$, so $\mathbf{V}$ and $\mathbf{V}'$ have the same eigenvalues. In the same way $\mathbf{J_2V} = \mathbf{WJ_2}$ and $\mathbf{V}$ and $\mathbf{W}$ have the same eigenvalues. Moreover we have that $\mathbf{V}'(\mathbf{0})$ has non-negative elements except on the leading diagonal. For $\sigma = 0$ the eigenvalues of $\mathbf{W}(\mathbf{0})$ are the same as those of $\mathbf{V}'(\mathbf{0})$. Furthermore if M is large enough $\mathbf{V}'(\mathbf{0}) + M\mathbf{I}$ is an irreducible matrix with non-negative elements and has an eigenvector $(\pi_1^*, \pi_2^*, \pi_3^*, \beta_1^*, \beta^*, \beta_{1+3}^*)$ which is strictly positive and has eigenvalue M. Hence this is the unique positive eigenvector and corresponds to a simple real eigenvalue which is also the spectral radius of $\mathbf{V}'(\mathbf{0}) + M\mathbf{I}$. Therefore all eigenvalues of $\mathbf{V}'(\mathbf{0}) + M\mathbf{I}$ lie in a circle centre the origin with radius M. Thus all eigenvalues of $\mathbf{V}'(\mathbf{0})$ and hence $\mathbf{W}(\mathbf{0})$ lie in the circle centre $(-M, 0)$ with radius M. Moreover zero is a simple eigenvalue of $\mathbf{W}(\mathbf{0})$.

The characteristic equation of $\mathbf{W}_1^+ = \mathbf{W}(0) - \sigma\mathbf{E}_1$, is of the form

$$\omega^6 + a_1(\sigma)\omega^5 + a_2(\sigma)\omega^4 + a_3(\sigma)\omega^3 + a_4(\sigma)\omega^2 + a_5(\sigma)\omega + a_6(\sigma) = 0, \qquad \text{(D.1)}$$

where $a_i(\sigma)$, for $i = 1, \ldots, 6$, are continuous functions of σ. When $\sigma = 0$ we have that $\mathbf{W}_1^+ = \mathbf{W}(0)$ and we already know that zero is an eigenvalue of $\mathbf{W}(0)$. Now consider the case where σ is small and positive, we have that the eigenvalues are continuous in σ in a neighbourhood about the origin. Hence for σ small we have that by continuity five eigenvalues will still have strictly negative real parts. We now examine the behaviour of the eigenvalue at the origin, say $\omega_6(\sigma)$, when σ is small and positive.

Lemma D.1 *If k_1, k_2, $k_3 > 0$ and $\hat{k} > 0$ are chosen appropriately then the constant term, $a_6(\sigma)$, in the characteristic equation of $\mathbf{W}_1^+$ is strictly increasing in σ for σ small and positive.*

Proof.

We wish to compute $\det \mathbf{W}_1^+$, this is equivalent to $\det \mathbf{V}$ when $\hat{\beta}, \pi_1, \pi_2, \pi_3$ are replaced by their respective lower bounds $\hat{k}\sigma$, $k_1\sigma$, $k_2\sigma$ and $k_3\sigma$. Denote this value of $\mathbf{V}$ by $\mathbf{V}_\sigma$. As $\mathbf{J_2}.\mathbf{V} = \mathbf{W}.\mathbf{J_2}$, $\mathbf{V}$ and $\mathbf{W}$ are reachable from one another using elementary row and column operations so $\det \mathbf{W}_1^+ = \det \mathbf{V}_\sigma$. Computing an expression for $a_6(\sigma)$ is straightforward but requires a large amount of algebra. Working along the second row of $\mathbf{V}_\sigma$ gives us two 5x5 determinants to compute, these in turn break down into the following 4x4 determinants:

$$\delta_1\lambda\gamma(\phi + \hat{\tau} + k_1\sigma(1-\phi)) \quad \times$$

$$\begin{vmatrix} -\lambda\hat{k}\sigma(1-\phi) & -\lambda\hat{k}\sigma(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\ \delta_2 & -(\mu+\delta_3) & 0 & 0 \\ \lambda\gamma(1-\beta^*(1-\phi)) & -\beta_2^*(1-\phi)\lambda\gamma & -\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi)) & -\lambda\gamma k_2\sigma(1-\phi) \\ 0 & \lambda\gamma(1-\beta_{1+3}^*(1-\phi)) & 0 & -\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \end{vmatrix}$$

$+$

$$(\mu+\delta_2)(\mu+\delta_3) \quad \times$$

$$\begin{vmatrix} -(\mu+\delta_1)-\lambda\hat{k}\sigma(1-\phi) & (1-\pi^*)\lambda\alpha_1(1-\phi) & (1-\pi^*)\lambda\alpha_2(1-\phi) & (1-\pi^*)\lambda\alpha_3(1-\phi) \\ \lambda\gamma(1-\beta_1^*(1-\phi)) & -\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi)) & 0 & 0 \\ -\beta_2^*(1-\phi)\lambda\gamma & -\lambda\gamma k_2\sigma(1-\phi) & -\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi)) & -\lambda\gamma k_2\sigma(1-\phi) \\ -\lambda\gamma\beta_3^*(1-\phi) & -\lambda\gamma k_3\sigma(1-\phi) & 0 & -\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \end{vmatrix}.$$

Note that we have denoted $k_1+k_2+k_3$ as k_{1+2+3} and k_1+k_3 as k_{1+3}. Expanding out these two determinants eventually gives us that $a_6(\sigma) = A+B+C+D+E+F+G$ where

$$A = -\delta_2\delta_1\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))\lambda\gamma(1-\beta^*_{1+3}(1-\phi)) \quad \times$$

$$\Big[(1-\pi^*)\lambda\alpha_3(1-\phi)\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi)) - (1-\pi^*)\lambda\alpha_2(1-\phi)\lambda\gamma k_2\sigma(1-\phi)\Big],$$

$$B = \delta_2\delta_1\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times$$

$$\Big[\lambda\hat{k}\sigma(1-\phi)\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi)) + (1-\pi^*)\lambda\alpha_2(1-\phi)\beta^*_2(1-\phi)\lambda\gamma\Big],$$

$$C = \delta_1\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))(\mu+\delta_3)\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times$$

$$\Big[\lambda\hat{k}\sigma(1-\phi)\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi) - \lambda\gamma(1-\beta^*(1-\phi))(1-\pi^*)\lambda\alpha_2(1-\phi)\Big],$$

$$D = (\mu+\delta_2)(\mu+\delta_3)\lambda\gamma(1-\beta^*_1(1-\phi))\lambda\gamma k_3\sigma(1-\phi) \quad \times$$

$$\Big[(1-\pi^*)\lambda\alpha_3(1-\phi)\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi) - (1-\pi^*)\lambda\alpha_2(1-\phi)k_2\sigma(1-\phi)\Big],$$

$$E = (\mu+\delta_2)(\mu+\delta_3)\lambda\gamma(1-\beta^*_1(1-\phi))\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times$$

$$\Big[(1-\pi^*)\lambda\alpha_2(1-\phi)\lambda\gamma k_2\sigma(1-\phi) - (1-\pi^*)\lambda\alpha_1(1-\phi)\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi))\Big],$$

$$F = (\mu+\delta_2)(\mu+\delta_3)\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))\beta^*_3(1-\phi)\lambda\gamma \quad \times$$

$$\Big[(1-\pi^*)\lambda\alpha_3(1-\phi)\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi)) - (1-\pi^*)\lambda\alpha_2(1-\phi)\lambda\gamma k_2\sigma(1-\phi)\Big],$$

and

$$G = (\mu+\delta_2)(\mu+\delta_3)\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times$$

$$\Big[(\mu+\delta_1+\lambda\hat{k}\sigma(1-\phi))\lambda\gamma(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi)) + \beta^*_2(1-\phi)\lambda\gamma(1-\pi^*)\lambda\alpha_2(1-\phi)\Big].$$

Unlike in the Simple Model and the Optimistic Models it is far from clear by examining the above expression whether $a_6(\sigma)$ is strictly increasing in σ. We have constructed the lower bounds such that we can choose k_1, k_2, k_3 and $\hat{k}$ to be any positive constants. Hence all we require to show is that $a_6(\sigma)$ is monotone increasing in any one of the four directions, $k_1\sigma$, $k_2\sigma$, $k_3\sigma$ or $\hat{k}\sigma$. Let $a_6(\sigma) = Y(k_1\sigma, k_2\sigma, k_3\sigma, \hat{k}\sigma)$. It is easy to identify

$\frac{\partial Y}{\partial(\hat{k}\sigma)}$ using the expressions $A-G$, hence we have that

$$\begin{aligned}
\frac{\partial Y}{\partial(\hat{k}\sigma)} &= \Big\{\delta_2\delta_1\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times \\
&\quad \left[\lambda^2\gamma(1-\phi)(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi))\right]\Big\} + \\
&\quad \Big\{\delta_1\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))(\mu+\delta_3)\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times \\
&\quad \left[\lambda^2\gamma(1-\phi)(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi))\right]\Big\} + \\
&\quad \Big\{(\mu+\delta_2)(\mu+\delta_3)\lambda\gamma(\phi+\hat{\tau}+k_1\sigma(1-\phi))\lambda\gamma(\phi+\hat{\tau}+k_{1+3}\sigma(1-\phi)) \quad \times \\
&\quad \left[\lambda^2\gamma(1-\phi)(\phi+\hat{\tau}+k_{1+2+3}\sigma(1-\phi))\right]\Big\} \quad > 0.
\end{aligned}$$

We wish to show that $Y(k_1\sigma, k_2\sigma, k_3\sigma, \hat{k}\sigma) > 0$ for σ small and positive. Using a Taylor series expansion about $(0,0,0,0)$ we have that

$$\begin{aligned}
Y(k_1\sigma, k_2\sigma, k_3\sigma, \hat{k}\sigma) &= Y(0,0,0,0) + \frac{\partial Y}{\partial(k_1\sigma)}k_1\sigma + \frac{\partial Y}{\partial(k_2\sigma)}k_2\sigma + \frac{\partial Y}{\partial(k_3\sigma)}k_3\sigma \\
&\quad + \frac{\partial Y}{\partial(\hat{k}\sigma)}\hat{k}\sigma. \qquad \text{(D.2)}
\end{aligned}$$

We have shown above that $\frac{\partial Y}{\partial(\hat{k}\sigma)}\hat{k}\sigma > 0$. By construction we can choose k_1, k_2, k_3 to be any positive values hence we can ensure that eqn (D.2) is strictly positive by choosing k_1, k_2, k_3 sufficiently small relative to $\hat{k}$. This completes the proof of Lemma D.1.•

Suppose that $\omega_6(\sigma)$ has a non-negative real part for σ small and positive, therefore $a_6(\sigma) = \omega_1(\sigma)\omega_2(\sigma)\omega_3(\sigma)\omega_4(\sigma)\omega_5(\sigma)\omega_6(\sigma) \leq 0$. However from Lemma D.1 this is impossible since we have that a_6 is strictly increasing in σ and $a_6(0) = 0$. Hence with our choices of $\hat{k}$, k_1, k_2 and k_3 for σ small and positive all eigenvalues of $\mathbf{W}_1^+$ must have strictly negative real parts.