MODELLING AIDS EPIDEMIOLOGY IN THE ERA OF HIGHLY ACTIVE ANTIRETROVIRAL THERAPY

B C Dangerfield and Y Fang

Centre for Operational Research & Applied Statistics

University of Salford

Salford M5 4WT

U.K.

Abstract

A system dynamics model of HIV/AIDS epidemiology is employed to simulate the effects of triple combination antiretroviral therapy in the treatment of HIV/AIDS. The epidemic data on homosexual men in the UK (1981–1998) have been used to fit the baseline model. The new combination therapies, which supplanted antiviral mono and dual therapy in 1996, are proving to be the most effective prophylaxis yet for halting viral replication *in vivo*, but many uncertainties still surround their use. A range of model-generated scenarios are created as a means of considering these uncertainties. Possible futures for the HIV/AIDS epidemic are constructed.

Keywords: system dynamics; optimisation; HIV/AIDS; health care; scenarios; planning

Introduction

Triple combination antiretroviral therapy (also named Highly Active Anti-Retroviral Therapy (HAART)) is currently the most appropriate treatment for advanced HIV disease and AIDS in those (mainly industrialised) countries which can afford it. It works by attacking HIV at different points in its life cycle so as to prevent it reproducing. Three classes of antiviral drugs are involved and the patient is required to take these at regular times throughout the day.

The introduction of this therapy, together with the viral load test, represents an encouraging milestone in the treatment of HIV/AIDS. Before the advent of the viral load test in clinics, both the health of people with HIV and the effect of treatments were assessed by monitoring changes in CD4 cell count, as well as physical signs and symptoms.

Prior to the introduction of HAART, no treatment regime had succeeded in satisfactorily stopping the progress of HIV disease *in vivo*. See for example the results of the Anglo-French Concorde trial in evaluating the effects of AZT as a monotherapy (Aboulker and Swart, 1993). This situation has changed greatly since HAART has been adopted. For many HIV-infected people, HAART can reduce the viral load in their blood to a level undetectable except by the most sensitive tests recently developed. Therefore, many HIV patients have been swapped from their original therapy over to HAART almost as soon as HAART was shown to be effective (Carpenter *et al*, 1996; Perelson *et al*, 1996; 1997). Its impact cannot be understated: one experienced American physician has described it as offering the greatest turnaround in prognosis that he has seen in 25 years in medicine (Carpenter, 1997). In the UK and other Western countries HAART was generally adopted around 1996.

The change in therapy not only changes the effects of treatment, but also influences the nature of the epidemic and the corresponding resource use. This will cause new and different impacts on the health care system and for planning purposes it is necessary to understand what these impacts might be. Published results on the effects of HAART (either medically or epidemiologically) are only just starting to appear (Gulick *et al*, 2000) but in any case it is important for national planning purposes to try to make headway now rather than await a thorough *ex post* analysis at a much later date. An important first step, therefore, is to offer likely futures for the HIV/AIDS epidemic in the era of HAART.

This paper is organised to consider issues surrounding the epidemiological consequences of improvements in survival time, knowledge about disease pathogenesis and uncertainty about the infectivity of individuals. Much of the model-based scenario work reported in this paper has been informed by very recent knowledge. Papers concerning the effects of HAART on individuals are starting to appear in the scientific and medical journals but our references lean heavily upon the proceedings of the (now annual) conference on retroviruses and opportunistic infections held in the USA.

HAART and the HIV epidemic

The influence of HAART on the HIV epidemic has both positive and negative aspects. On the one hand, the therapy is effective in delaying or halting disease progression. Logically, this will have two consequences. Firstly, the number of diagnosed AIDS cases will exhibit a sharp decrease consequent upon the widespread

QUARTERLY DIAGNOSED CASES OF AIDS IN THE UK (HOMO/BISEXUAL MEN)
Final 12 quarters adjusted by the method of Heisterkamp et al (1989)
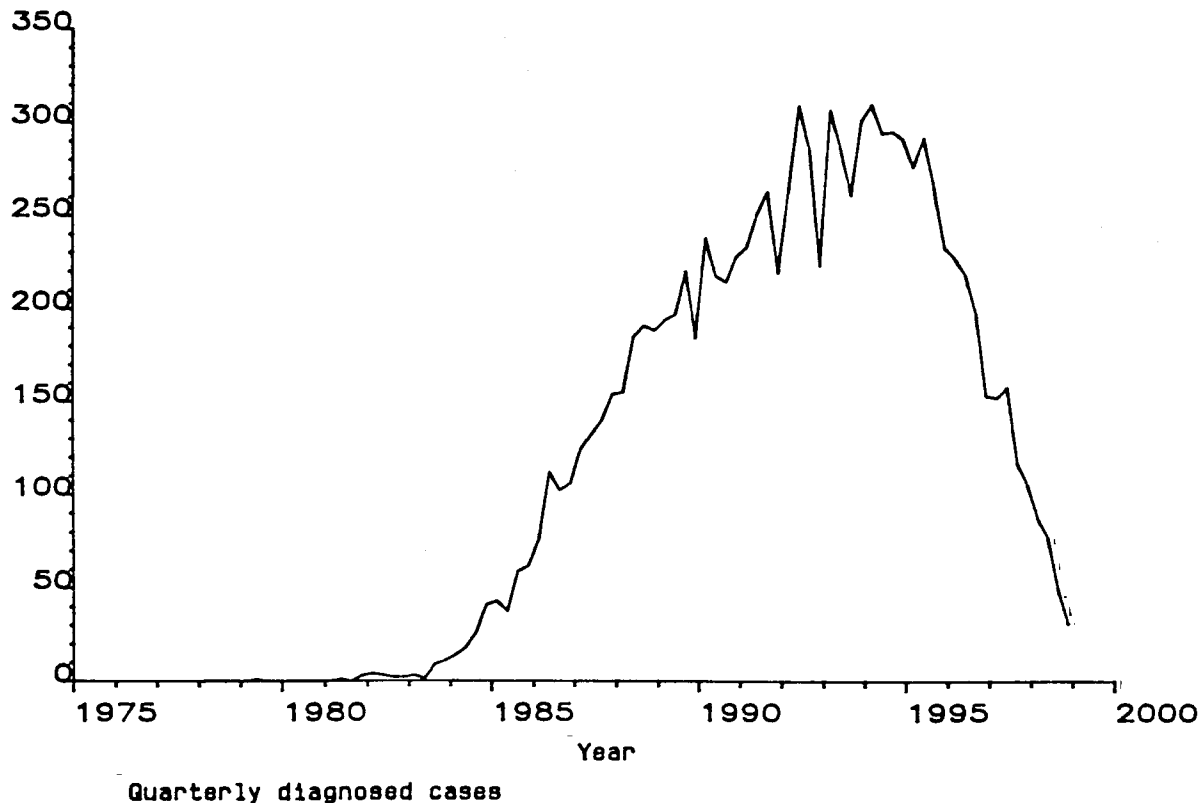
Quarterly diagnosed cases

Figure 1: *Sharp decline in the diagnosed incidence of AIDS in the UK consequent upon the administration of HAART in 1995/96*

adoption of HAART. The data on new AIDS cases in the UK in recent years following the introduction of HAART has demonstrated this (see figure 1). Similar pictures prevail in other western developed societies which have adopted HAART.

A second consequence is that, logically, the prevalence of HIV/AIDS will increase. This will have serious repercussions for future HIV incidence if the infectivity of seroprevalent persons stays at the same level. However, the new therapy has been successful in achieving a substantial viral load suppression and this makes for the more optimistic consequence that the infectivity of a person undergoing HAART is, in all probability, much lower than that of an infected person not receiving it. Therefore, the widespread adoption of HAART is likely to result in a lower HIV incidence and, in the long run, it will decrease HIV/AIDS prevalence. This belief cannot, as yet, be confirmed but it is an opinion which is gaining increasing prominence and it is a central aspect of one of our simulated scenarios.

The model adopted in this study

The system dynamics model of HIV/AIDS epidemiology employed in this study has developed over some twelve years, undergoing enhancements and changes as the flood of knowledge about HIV and AIDS has grown (Roberts and Dangerfield, 1990; Dangerfield and Roberts, 1994; 1996b). It has now been adapted to offer a best fit of the model to data reflecting the seemingly highly encouraging effects of HAART on recent AIDS incidence. In this way, not only is a reference mode made available from which a series of scenarios about the future epidemiological impact of HAART can be developed, but the HAART parameter estimates from the fitting process constitute interesting information in their own right. Figure 2 gives the epidemiological sector of the model prior to the introduction of HAART therapy.

The sector shows a three stage incubation period which corresponds with an assumed high-low-high profile of infectiousness. The 'on treatment' component here reflects the previous antiviral monotherapy (Hoover, 1995; Delmas *et al*, 1995). The AIDS stage is split into early and late stages, the latter reflecting the final few months of care when symptoms are likely to be very pronounced and the patient is most likely to be hospitalised.

It is important to emphasise that, although not shown on the flow diagram, HIV infection and the subsequent incubation period is replicated to reflect three strata of susceptibles and infecteds each with differing levels of sexual activity and proportionate mixing between and within these strata. Finally, there is a flow out of each compartment to allow for non AIDS-related deaths; the inflow of new susceptibles exactly balances this loss.

The model structure needs to be adjusted to cope with the implementation of HAART and to provide a new model variable which depicts the numbers of patients on HAART at any one time, for this is crucial in terms of the costs of the therapy. Figure 3 (a and b) offers two possible model structures, called models 1 and 2.
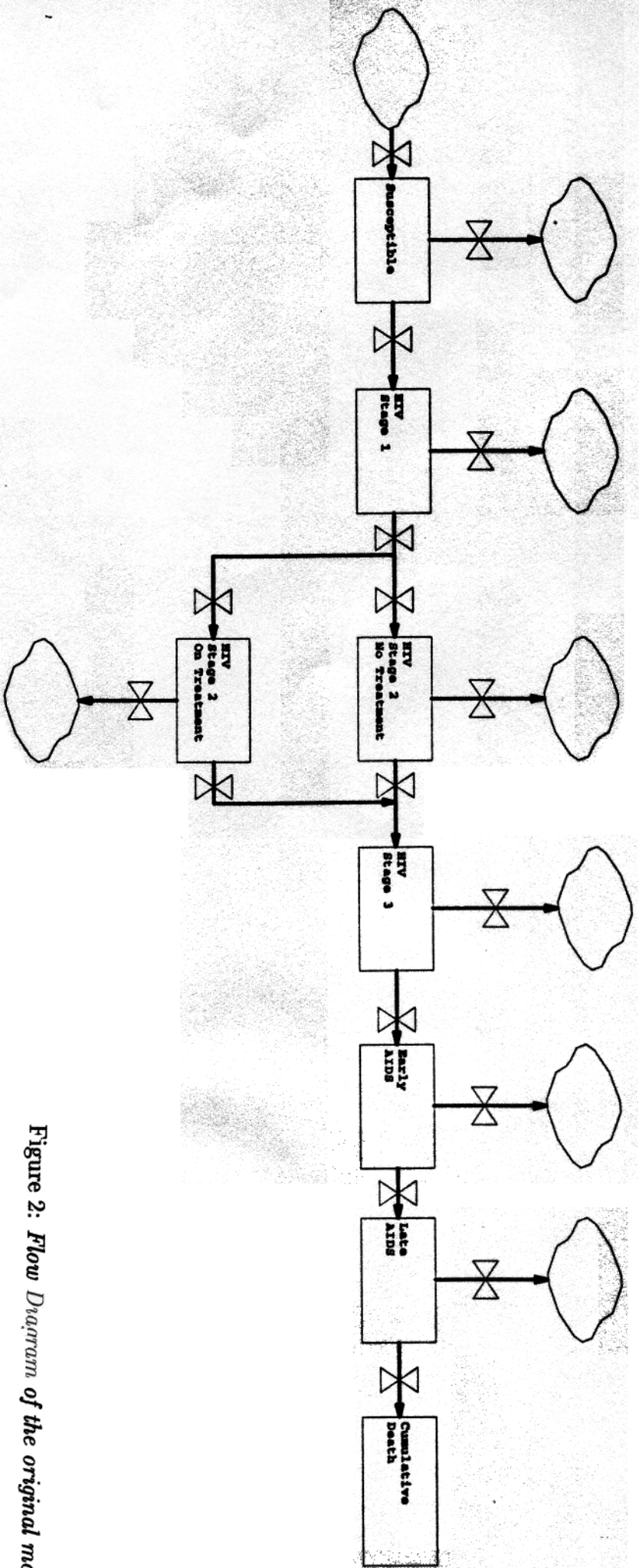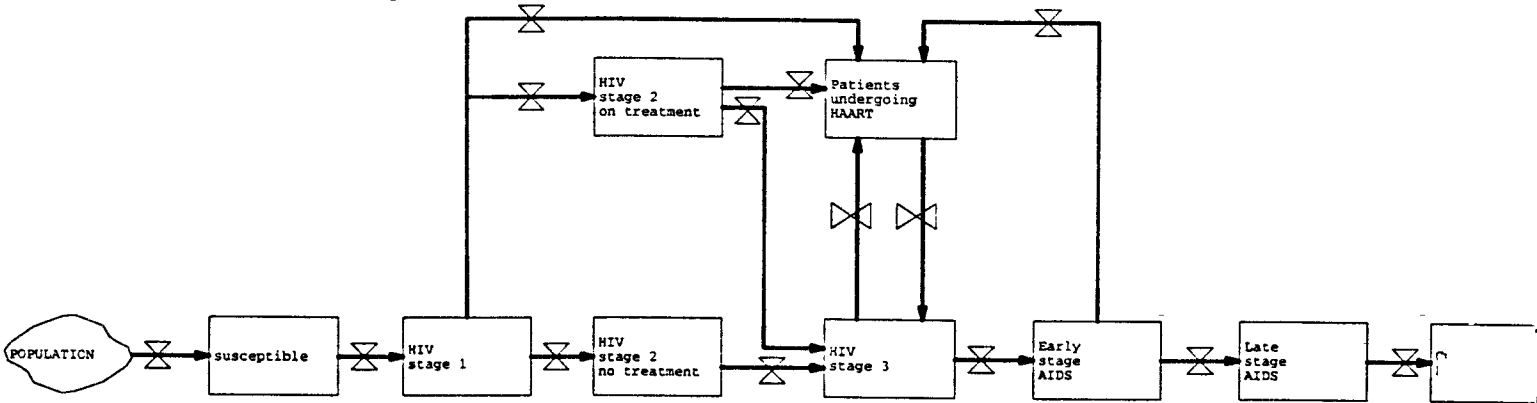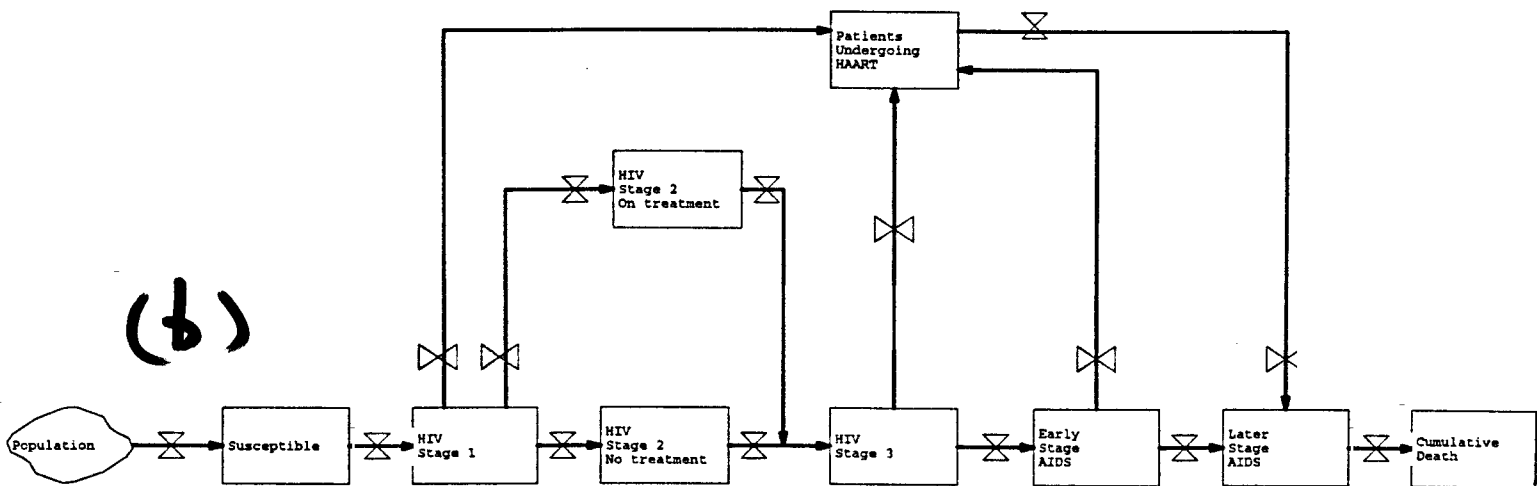
Figure 2: *Flow Diagram of the original model*

Figure 3: *Flow diagrams of the first two HAART models: (a) therapy breakdown results in a flow of cases to advanced HIV disease (Model 1) and (b) to early-stage AIDS (Model 2)*

These two are differentiated only insofar as the flow of cases in the event of therapy breakdown results in a move to the state of either advanced HIV disease or late-stage AIDS. Given that some patients on HAART emanated from confirmed AIDS cases it might seem illogical for these to then progress to advanced HIV disease. The surveillance data certainly do not account for this eventuality – once a person is classified as having AIDS that does not change regardless of any (or vast) clinical improvement in their condition (Downs, 1999).

No flow is shown from patients at HIV stage 2, but not on treatment, direct to the state of being on HAART. It is assumed that such cases are unaware of their serostatus until they reach HIV stage 3. Also, the outflows due to non-AIDS deaths and shown in figure 2 are not repeated, for convenience.

Another model structure (model 3) which we have explored involves separating out patients on HAART according to their disease state at the commencement of therapy. It may well be that HAART is more effective the earlier the stage of viral pathogenesis in the host at the commencement of therapy. Also there is the possibility that patients on previous mono or dual antiretroviral therapy will be better placed to assimilate the new regime than those who have never before been given such therapy. The flow diagram for this model is shown in figure 4. Again the outflows due to non-AIDS deaths, although in the model, are not shown in the flow diagram.

With the flows as depicted in the diagrams, the model becomes considerably more complex than the corresponding model prior to the introduction of the new treatment state (Dangerfield and Roberts, 1994). For instance, it is important to ensure that the models keep track of all cases. An elaborate system of mass balance equations is embedded in each model to ensure that there is no unintentional gain or loss in the number of cases. Further, prior to the calendar time of the implementation of HAART (1995/96) no flow into the HAART state is allowed to take place and, more importantly, any flow out of HAART following therapy breakdown is unidirectional: no recycling back into HAART is permitted.

Fitting the baseline models

All three of the above model structures have been fitted to the data on diagnosed AIDS cases in the UK depicted in figure 1. The methodology adopted is heuristic search in which the parameter space is explored using a hill-climbing algorithm and repetitive runs of the model (Dangerfield and Roberts, 1996a).

The optimisations in this case involved a temporal pivot. Prior to 1995-96 the set of parameters optimised consisted solely of those optimised in previous exercises (Roberts and Dangerfield, 1992; Dangerfield and Roberts, 1996b). Following the introduction of HAART, however, an additional set of parameters, peculiar to the new therapy, become active. There is thus a two-stage optimisation procedure.

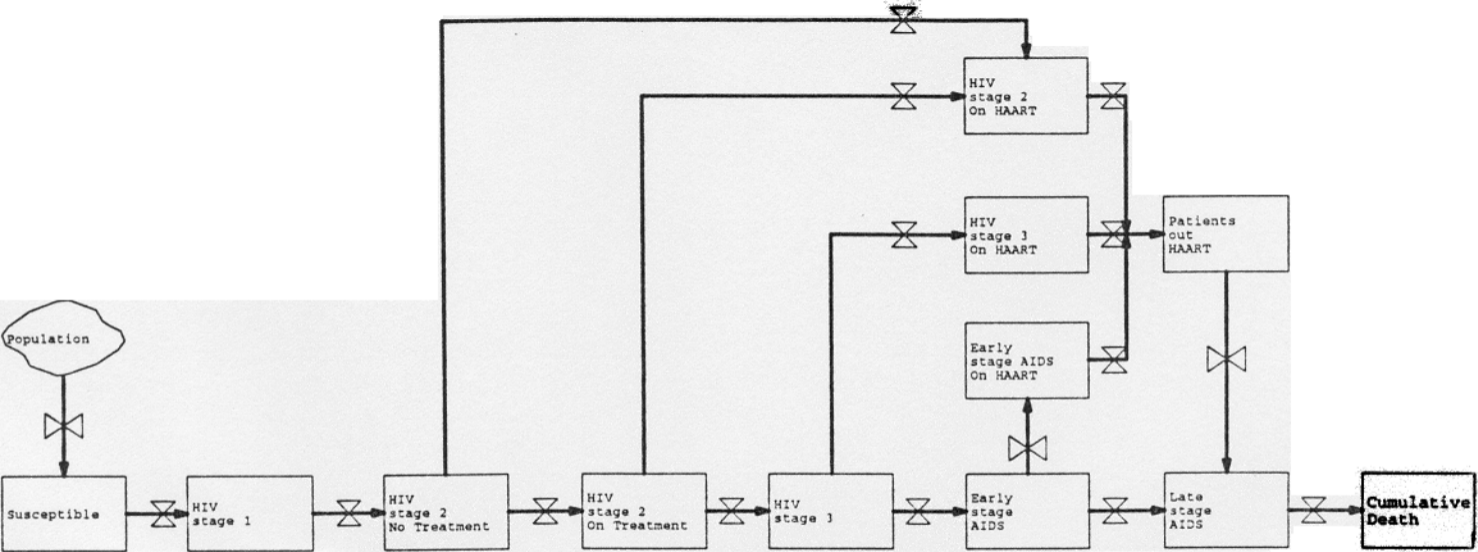The optimised parameter values are given in table 1.

Figure 4: *Flow diagram of the third HAART model: therapy is differentiated according to the disease state reached prior to assimilation on HAART*

Table 1: Optimised parameter values for each of the three models (parameters below the double horizontal line relate to triple therapy only; those starred [*] are set outside of the model fit; all times are in years)

PARAMETER	MODEL 1	MODEL 2	MODEL 3
*STIMEBC Start TIME of Behaviour Change	1984	1984	1984
*STIMEAZT Start TIME of AZT monotherapy	1988	1988	1988
*MSTC Mean Survival Time after AIDS, Constant	0.9	0.9	0.9
*MSTFI Mean Survival Time after AIDS, 1st Increase	0.9	0.9	0.9
PH1ITC PHase 1 Incubation Time, Constant	0.374	0.374	0.376
PH2ITC PHase 2 Incubation Time, Constant	8.956	8.956	8.071
PH2ITFI PHase 2 Incubation Time, First Increase	0.5	0.5	0.48
PH3ITC PHase 3 Incubation Time, Constant	1.741	1.741	1.422
PH3ITFI PHase 3 Incubation Time, First Increase	0.276	0.22	0.22
*DLSAIDS Duration of Late Stage AIDS	0.25	0.25	0.25
AVRT AVerage Residence Time in the population	25.4	25.4	25.4
ETPS1AR Estimated Total Population of Stratum 1 At Risk	366904	366904	371575

```
   ETPS2AR                          208486     208486     210746
Estimated Total Population of Stratum 2 At Risk

   ETPS3AR                          110296     110296     110296
Estimated Total Population of Stratum 3 At Risk

   PIPP1                            0.125      0.125      0.126
Prob. of Infection Per Partner, stage 1

   PIPP3                            0.0174     0.0174     0.0203
Prob. of Infection Per Partner, stage 3

   CRAT                             0.0232     0.0232     0.0229
Constant RATio relating PIPP1 to PIPP2

   ALPHAS1                          1.16       1.16       1.12
   ALPHAS2                          7.15       7.15       7.36
   ALPHAS3                          36.03      36.03      36.04
Coefficients for partner numbers functions, strata 1-3

   BETA                             3.200      3.200      2.737
Used in the partner numbers function

   GAMAS1                           0.524      0.524      0.591
   GAMAS2                           0.524      0.524      0.447
   GAMAS3                           0.523      0.523      0.525
Coefficients of decrease in partner numbers, strata 1-3

   ZETA                             1.073      1.073      0.865
Power coefficient in the behaviour change function
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------

   STIMETRI                         1995.5     1995.5     1995.25
Start TIME of TRIple therapy

   TTST                             15.74      20
Triple Therapy Sojourn Time

   MIITTC                                                 11.5
Mean Increase in Incubation time due to Triple Therapy, Constant
```

```
    PERTT1C                                    0.68        0.60
% diverting to Triple Therapy at stage 1, Constant

    PERTT2C                                    0.64        0.53
% diverting to Triple Therapy at stage 2, Constant

    PERTT3C                                    0.1         0.25
% diverting to Triple Therapy at stage 3, Constant

    PERTTEAC                                   0.1         0.35
% diverting to Triple Therapy from Early-stage Aids, Constant

    PERTT2NC                                               0.60
% diverting to Triple Therapy at stage 2; Not yet on treatment, Constant

    PERTT2UC                                               0.57
% diverting to Triple Therapy at stage 2; Under treatment, Constant

    COTIME                                     2.32       2.53       1.66
    COPOWER                                    4.99       5.00       1.0
COefficients of function relating to % diverted onto triple therapy from
    various states; gives rise to the following maxima at end-1998...
(Max % on triple therapy)                     66.4       65.3       67.2

    MLOCT                                      2.67       2.56       2.37
Mean Length Of time to Change to triple Therapy (mean of an exponential)

    LOFCT                                      3.0        3.14       3.25
Length OF Change Time to triple therapy (truncation time)

    PIPPT                                      0.5E-03   0.5E-03    0.5E-03
Prob. Of Infection Per Partner when on Triple therapy
```

Changes in survival and mean incubation times prior to HAART

Irrespective of the apparent benefits of HAART, earlier treatment advances had ensured that the average survival time, consequent upon an AIDS diagnosis, and also the mean incubation time lengthened with calendar time. To reflect these changes, the mean of the exponential distribution governing the two instances has been extended according to the following function of calendar time.

$$f(t) = a + b(1 - EXP(-t^2)) \text{ where}$$

a = original mean survival time (MSTC)
b = 'first' increase in mean survival time prior to advent of HAART (MSTFI)
t = time from introduction of monotherapy

The functional form employed ensures that the mean changes according to a smoother transition than would be obtained via a step function: for the survival time the mean changes from 0.9 yr to 1.8 yr consequent upon the introduction of AZT in 1988.

For the incubation period of HIV the second and third phases of the virological staging model are increased as from 1988: in the case of models 1 and 2 the average rises from 8.96 yr by 0.5 yr in phase 2 and from 1.74 yr by 0.28 yr and 0.22 yr respectively.

Infectivity

The three-stages of HIV pathogenesis are associated with different levels of infectivity of the host. A high probability is attached to the first phase (PIPP1). This is related to the level in the second (asymptomatic) phase by a fixed ratio (CRAT) which produces an infectivity probability in the second phase of 0.0029 for all three models. The third phase sees a rise to 0.017 (PIPP3). As in the case of earlier model fits (Dangerfield and Roberts, 1994; 1996b) the U-shape of infectivity is reconfirmed.

Reduction in number of new partners

A functional form has been used in recent versions of the model to account for changing sexual behaviour. This has replaced the use of a table function which was felt to be more arbitrary and increased the number of parameters to be optimised. The function below has four parameters whilst the table required the optimisation of five (even in a coarse form) and could often produce an irregular curve.

$$f(t) = \alpha((1 - \gamma) + \gamma EXP(-(t/\beta)^\zeta))$$

The time, t, for the start of behaviour change is determined to be 1984 which is consistent with known data from sociological surveys and the general increase in prominence of the threat of AIDS, together with associated health education campaigns at that time.

Strictly there is an identifiability problem in estimating both the infectivity probability and the partnership formation parameters simultaneously. However, with our model formulation there exists a parameter set which is a function of calendar time (partnership formation) taken together with another trio of parameters which are a function of biological time (infectivity). It is a testimony to the system dynamics approach that such a conjunction is straightforward; mathematical methods would

find it rendered the model rather intractable. The identifiability problem is acknowledged, but it is a fact that a less good fit to the data is obtained with a model incorporating a constant infectivity probability over the entire incubation period.

HAART parameters

The initial time of introduction of HAART

The HAART clinical trial began in 1995 and the literature which provided the first recommendations as to the clinical practice of HAART was published in 1996 (Carpenter *et al*, 1996). These recommendations have been updated annually (Carpenter et al, 1997; 1998). In 1996, HAART gradually came into common use in industrialised countries. However, our results revealed that mid-year 1995 gave the best fit start time, which is consistent with physicians assimilating many patients onto HAART as part of clinical trials in the UK.

Flows into and from HAART

Most HIV patients in the industrialised world are offered HAART when they are confirmed as seropositive. For example, New York patients have begun treatment on average some 120 days after the presumed date of primary infection (Blakeslee, 1999). This would typically correspond to the end of the phase of acute infection and the beginning of the asymptomatic stage. However, in 1996 a large percentage of patients already on monotherapy were transferred onto HAART treatment. It is reasonable to assume that no treatment of any kind is possible if the infected person is not aware of their infection and that at late-stage AIDS it is too late for even HAART to be effective. Therefore, we have chosen to model patients entering onto a regime of HAART from HIV stage two, three and from early-stage AIDS.

In models 1 and 2, the sojourn time in triple therapy is estimated from the data fit. It will be recalled that the two models provide for a transfer to advanced HIV disease and early-stage AIDS respectively in the event of therapy breakdown. The fits reveal a sojourn time of 15.7 yr and 20 yr respectively from the currently available data. These figures have to be compared with the duration of time patients would have been incubating HIV if HAART never became available. They are also minima. Increases in the sojourn time have an almost imperceptible effect on the fit and, with the release of additional data showing a continuing decline in AIDS incidence, these values can be expected to increase. A reduction in the sojourn time produces an earlier bottoming out of the AIDS incidence curve.

Model 3 has a different structure to the other two models. Whereas in models 1 and 2 the estimated duration on triple therapy is, in part, parallel with the other possible states, for model 3 the transition to triple therapy *follows* a period in one or other state. Thus the mean increase in incubation time of 11.5 yr has to be added to mean duration times in previous states in order to be comparable to the figures

of 15.7 and 20 yr respectively for the other two plausible structures.

The parameters PERTTxxx relate to the parameters used to bifurcate the flows emerging from various HIV states and early-stage AIDS onto HAART. It has been assumed that no transfers from late-stage AIDS to HAART were made. Higher values of these proportions were obtained for the earlier stages of HIV pathogenesis.

Across all three models, the estimated maximum percentage of all HIV and AIDS cases on HAART in the UK had, by end-1998, reached $\approx 66\%$. Estimates provided by Beck et al (1999) for the UK show 51% of symptomatic HIV cases were prescribed the therapy in 1997 whilst the estimate for AIDS cases was 69%. In the USA, a study conducted in 1997 by Markson (1999) collected data by interviewing 2267 patients all of whom had CD4 cell counts below 500 and a viral load greater than 10,000 copies/ml. The study found that two-thirds of the patients were receiving HAART. Our result is not inconsistent with these other estimates.

It is assumed that once triple therapy was acknowledged to be efficacious, physicians transferred patients onto it reasonably quickly. The transfer rate, from the various possible states, is modelled by a first-order exponential delay. However, the rate is truncated since otherwise there would be a long right-hand tail of transfers which is considered to be unrealistic: physicians would wish for their patients to receive the benefits of the new therapy as rapidly as possible. The parameters MLOCT and LOFCT represent the mean and (truncated) limit of transfer time, respectively. This rapid assimilation of patients onto HAART stores up a potential problem later should HIV manage to defeat the intentions of the therapy. Part of the scenario analysis described below investigates this possibility.

Estimated infectivity of those on HAART

The infectivity of patients on HAART is estimated to be very low: a value of 0.0005 was obtained for all three models. There is considerable uncertainty about this aspect amongst the medical profession and so it is a prime candidate for consideration in the scenario analysis.

Model fits

The chi-square values for the three models were 90.3, 90.6 and 89.5 respectively. By a narrow margin, model 3 offers the best fit. Model 1 is used for the scenario explorations described later and figure 5 shows that, with the exception of four outliers discussed below, all the data points are contained within upper and lower bounds to the mean estimate generated by this model and given by $\mu \pm 2\sigma$ where $\sigma = \sqrt{\mu}$ since the rate of diagnosis is assumed to be a Poisson process.

EPIDEMIOLOGICAL SCENARIOS FOR HAART IN THE U.K.
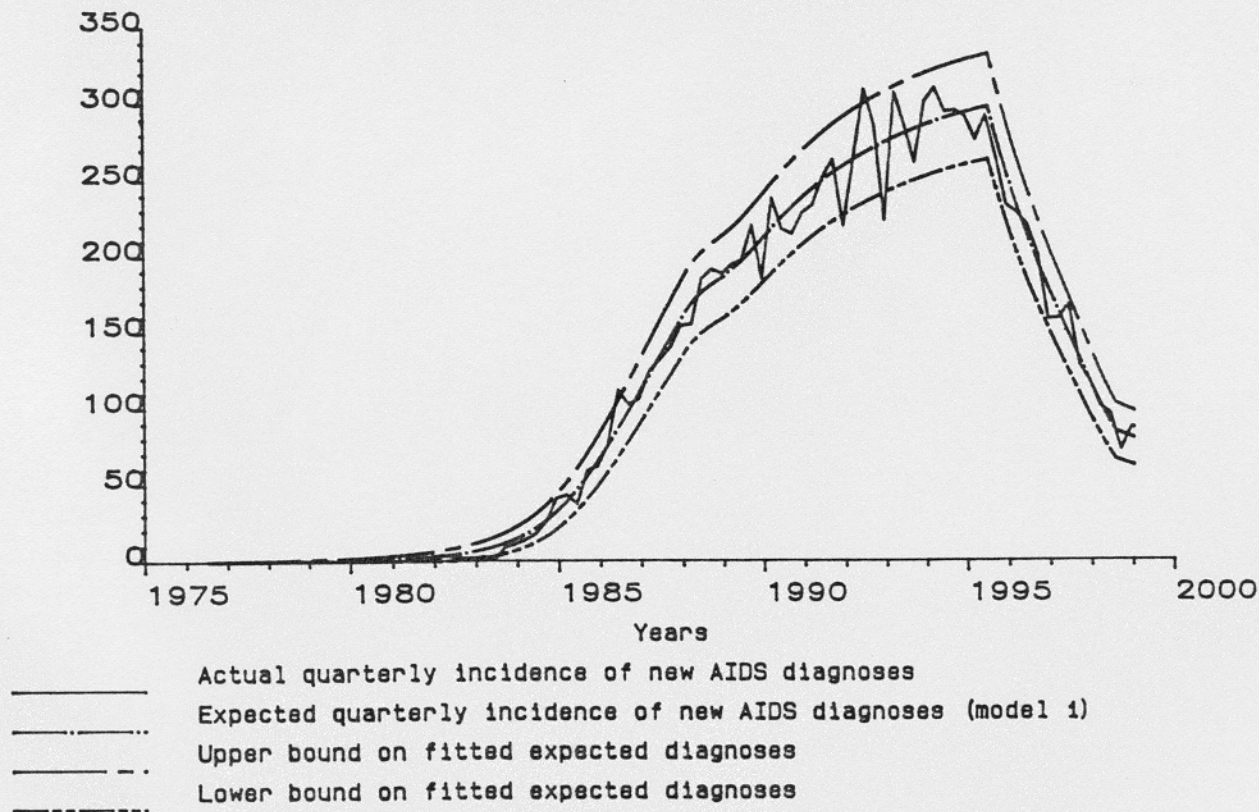Model generated bounds on quarterly incidence of new AIDS diagnoses

Actual quarterly incidence of new AIDS diagnoses
Expected quarterly incidence of new AIDS diagnoses (model 1)
Upper bound on fitted expected diagnoses
Lower bound on fitted expected diagnoses

Figure 5: *Expected and observed values of new quarterly diagnosed cases of AIDS in the UK together with bounds on the central estimate (model 1)*

Model-generated scenarios

The scenario results reported below stem from work with the model 1 structure (figure 3(a)) but, except where indicated, the same qualitative conclusions arise through explorations with the model 2 and model 3 structures also (figures 3(b) and 4, respectively).

The effects of HAART on survival time and infectivity

For the purpose of projecting future trends in the HIV epidemic, certain aspects of the effects of HAART need to be considered. The first of these centres upon its effects on viral pathogenesis. The effects of HAART are measured both by a viral load test and CD4 cell count. A number of studies (Perelson *et al*, 1997; Lewin *et al*, 1999; Gulick et al, 2000) show that, for many infected people, HAART has reduced the number of viral particles in their blood to a level undetectable except by the latest technology, although a mathematical model of CD4 cell dynamics, consequent upon HAART and fitted to patient data, suggests there is still residual HIV replication (Ferguson *et al*, 1999). The reduction in viral load is usually accompanied by the disappearance of symptoms and the restoration of weight previously lost. The onset of opportunistic infections (and the need to treat such symptoms) thereby disappears.

Unfortunately, it is not possible to tell how long the efficaciousness of HAART can be maintained. With currently available medical knowledge it is uncertain whether HAART can halt disease progression forever (provided the regime is strictly followed), or whether it will delay disease progression only for a limited, but perhaps relatively lengthy, duration. Because HIV has the ability to mutate quite easily, there is no certainty that current success with HAART will be maintained forever. It may well be that current achievements are merely transient but with a longer duration of efficaciousness than when monotherapy was first adopted.

With current knowledge it is thought most likely that HAART is a treatment which must last for a very long time. Although Perelson *et al.* (1997) and Gulick *et al.* (2000) have indicated that HIV can actually be eradicted after about 3 years of complete virus suppression, other medical researchers are not so sure. (See for example Wong *et al.* (1997)).

It is possible to discern two possible outflows from the state of being on HAART: one is the death of a patient for a reason unconnected with AIDS and the other is treatment failure. In the model a negative exponential function is employed to describe the residence lifetime of the patient population so that non-AIDS deaths can be taken account of. However, modelling the outflow caused by patients experiencing treatment breakdown encounters two problems:

- How long is the average duration of efficaciousness of HAART in a typical patient?

• Which time profile is suitable to describe the distribution of the sojourn time under efficacious treatment?

At the moment it is not possible to state categorically that patients faithfully maintaining their HAART regimen will never develop AIDS. Although the early optimism surrounding use of AZT as a monotherapy in the late 1980's was eventually shown to be misplaced, there are some clinicians who think that, this time, HIV can really be kept in check. HIV has proved to be a very cunning virus to combat thus far and its ability to overcome even therapy as seemingly effective as HAART must be considered. Overlaying this is the additional uncertainty of the medical effects of long-term use of antiretrovirals.

Accordingly, in the scenarios described below we have sought to explore, as a primary objective, the possible effects of limited efficaciousness with HAART. We have considered a range of durations for efficacious treatment of 10 yr, 15 yr and 20 yr; these durations correspond to sojourn times under HAART.

Another unresolved issue is the level of infectivity of people undergoing therapy with HAART. Theoretically, HIV infected people with a lower viral load should exhibit lower infectivity. However, there is currently only fragmentary evidence (Andreoni *et al*, 1997; Wolinsky, 1998; Vernazza *et al*, 1997) showing that a decrease in viral load does correlate with a decreased risk of transmitting HIV. Some of the early evidence comes from a study of 39 seropositive hæmophiliacs, all male, and their HIV-negative sexual partners. Researchers from the Hæmophiliac Centre of Western Pennsylvania at Pittsburgh found that men with higher viral loads were significantly more likely to transmit HIV to their partners during unprotected sexual intercourse. However, more recently Cohen (1998) and Coombs (1998) show that there is still some work which needs to be done to answer this question definitively.

The efficacious period for HAART and infectivity: their impact on new HIV infections

It is appropriate to consider the effects of variations in the efficacious period for HAART alongside variations in infectivity whilst on the new therapy. Figure 6 allows a contrast between the scenarios of changes in each insofar as they impact on new cases of HIV infection. The figure is one of enlarged detail from 1992 when, in the absence of HAART, the trajectory of new infections was firmly on the decline anyway. But widespread administration of HAART from 1995/96 causes a sharp and dramatic reduction in HIV incidence based on the assumption that infectivity whilst on the new therapy is practically zero. In fact the probability is set to 0.0005 for the baseline efficacious period of 15.74 years.

What is particularly striking about the contrast between figure 6(a) and (b) is that the changes to the average efficacious period have only a modest impact on new infections for a given infectivity probability (figure 6(a)). The slight rebound, most
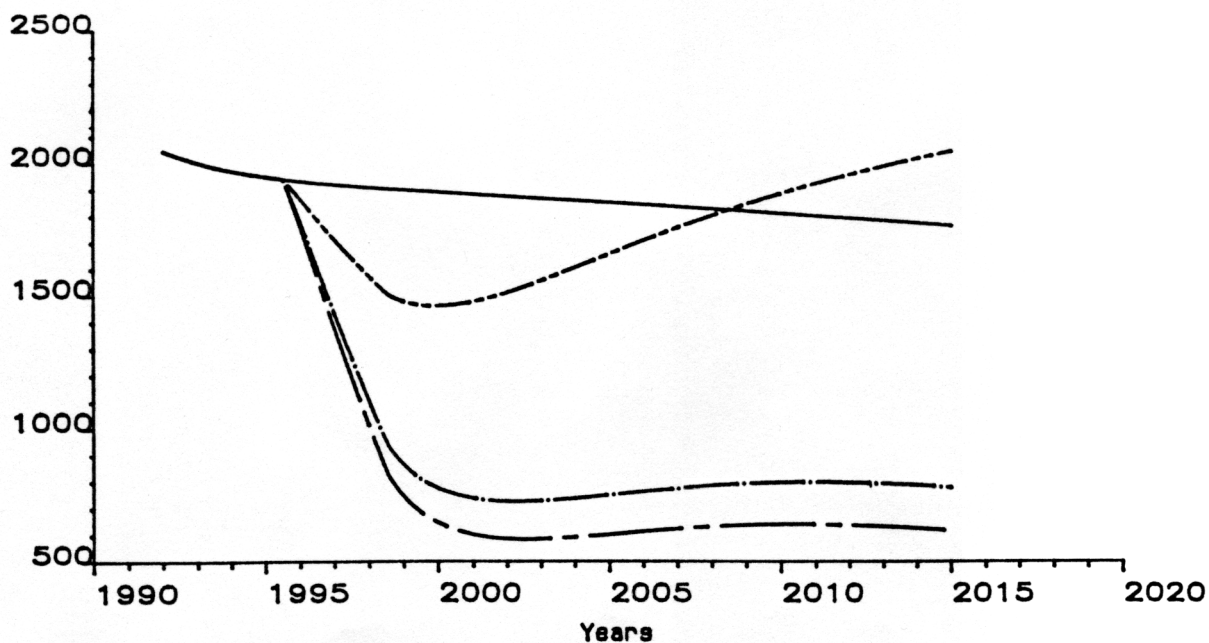
Figure 6: *Enlarged detail plot of the effect on HIV incidence of (a) variations in three assumed average efficacious periods for HAART and (b) variations in the infectivity probability when on HAART*

evident in the scenario for an efficacious period of 10 years, is explained by cases emerging from triple therapy relatively quickly, entering HIV stage 3 (by assumption) and therefore able to infect others. With model 2 (figure 3(b)) the assumption is that such a breakdown in therapy results in patients transferring to late-stage AIDS where they have no physical capability to infect others. In that case the three curves in figure 6(a) are extremely close together.

By contrast, as infectivity is both reduced and increased by five times, a rather dramatic change in new infections is observed. For the largest infectivity probability the increase has a greater than proportionate effect.

This result has profound implications. If there is a residue of doubt about the infectivity of a patient on HAART who continues (or commences) to practice unsafe sex, then any slackening of health education campaigns could have dire consequences. Fortunately the balance of the evidence seems to suggest that a reduction in viral load offers a concomitant reduction in infectivity, but this assertion has not yet been proven. For patients on HAART to believe that use of a condom is no longer necessary might be the kind of reaction which severely undermines the beneficial effects of the therapy.

The possible consequences of treatment breakdown: its impact on new AIDS cases

The choice of the time profile to model treatment breakdown is obviously clouded by uncertainty and this too can be dealt with by generating separate scenarios. It might be thought that the statistical hazard of treatment breakdown is constant, implying a first-order delay effect (negative exponential) from the level representing the numbers on HAART. However, this also implies, based upon a consideration of the transient behaviour of material delays of various orders (see for example Sterman (2000)), that the greatest rate of treatment breakdown follows directly from the initial administration of HAART. Around 1996 a substantial number were transferred to HAART (if not already on it) and there is no evidence of any significant breakdown yet, given three or four years' worth of clinical experience with the therapy. It must be concluded that a first-order delay effect is unlikely.

The transient behaviour following the introduction of HAART in a group of patients might evolve as an initial period of widespread benefit followed by increasing numbers experiencing treatment breakdown. The Erlang3 distribution exhibits such a profile and also has a hazard function which continuously increases at an decreasing rate and would thus seem to be a more appropriate choice, implying a third-order material delay effect.

Following treatment breakdown, should HIV overcome HAART, the model structure issue which must be addressed is to which state (level) the patients move to. Unless an as yet undiscovered treatment becomes available, the two possibilities which present themselves are those mentioned earlier: one is that patients move to HIV stage 3 infection whilst the alternative is that they move direct to late-stage

AIDS. Neither one of these possibilities can be conclusively rejected on the basis of currently available knowledge.

The effects of varying the parameter related to the length of the efficacious period for the new therapy are now considered in respect of the likely impact on new AIDS cases. It is a parameter representing the average residence time in triple therapy and, as suggested above, is consistent with a third-order Erlang distribution. Whether this is a reasonable assumption will be revealed over the coming years, perhaps decades. Figure 7 charts the effects on new cases of AIDS

The dramatic reduction in new AIDS cases consistent with the data fit, now experiences a rebound effect, the sharper and more immediate the lower is the average duration of the efficacious period. Consideration of the time response of material delay structures makes this an unsurprising result. In the 1995-96 period, the rapid assimilation of new patients onto HAART is almost equivalent to a pulse, although the reduction in the HAART assimilation rate would not be completely symmetrical to the rise given additional HIV/AIDS cases coming forward for testing and treatment.

An associated factor here is the period of time over which physicians transferred their patients onto HAART. Five average durations: 0.5, 1, 1.5, 2 and 2.5 yr have been considered and the joint effect of this change simultaneous with the changes in the average duration of the efficacious period can be seen in figure 8. Here AIDS prevalence is used since this is more meaningful in contemplation of resource provision.
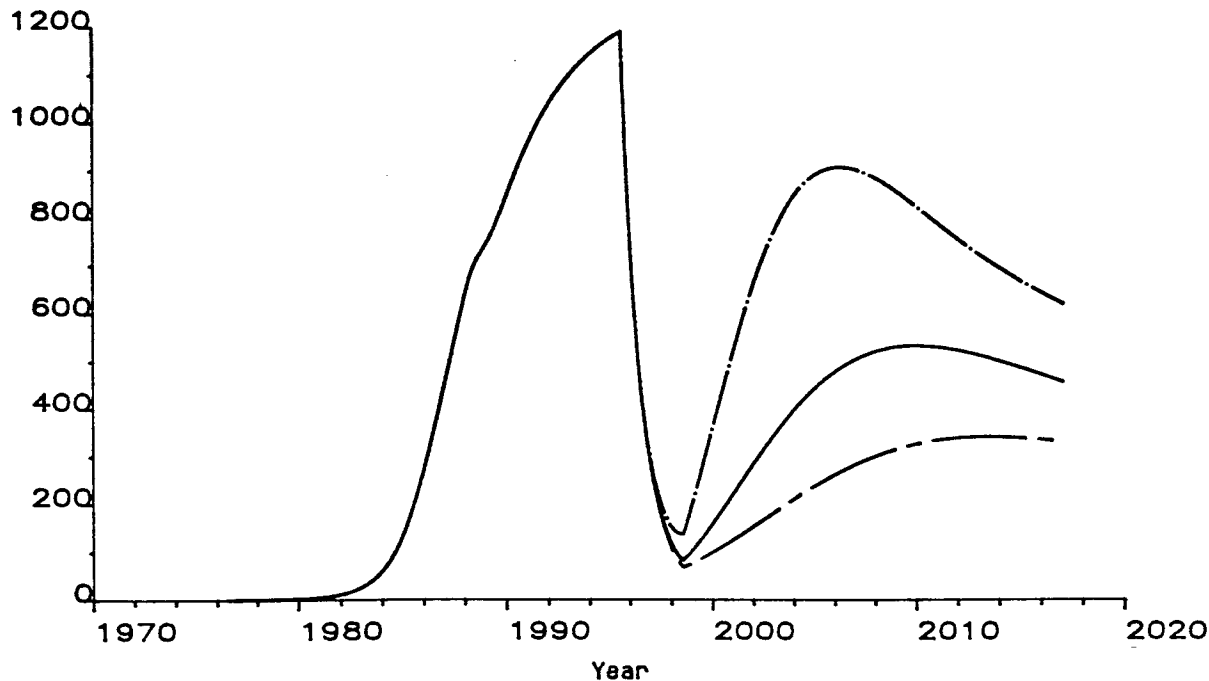
The shorter is the HAART assimilation period and the lower the assumed average efficacious period, the higher is the peak of prevalent AIDS cases when the numbers rebound. This effect is more pronounced the lower is the average period of efficaciousness. Some consolation is offered in that the data fitting discussed earlier suggests that the assimilation period is slightly larger than the 2.5 years assumed to be the longest period of assimilation for this scenario analysis.

For resource provision the above result provides a salutory warning. The advent of HAART has seen AIDS wards in hospitals being closed as the number of cases with severe symptoms have dropped sharply. This rebound effect, associated with a large number of patients being assimilated onto HAART over a reasonably short period, would imply that the health service will be faced with a sharp increase in patients who were quite ill and in need of hospitalisation. A rapid implementation of resource provision would inevitably be required.

Regression in sexual behaviour by those on HAART

An obvious concern in the health care system emerges when patients with HIV/AIDS, together with those currently uninfected but at risk, come to appreciate the changed prognosis in terms of their future well-being. There is an obvious temptation, especially on the part of new and younger male homosexuals, to ignore exhortations towards safe sex and to revert to practices evident amongst their predecessors in co-
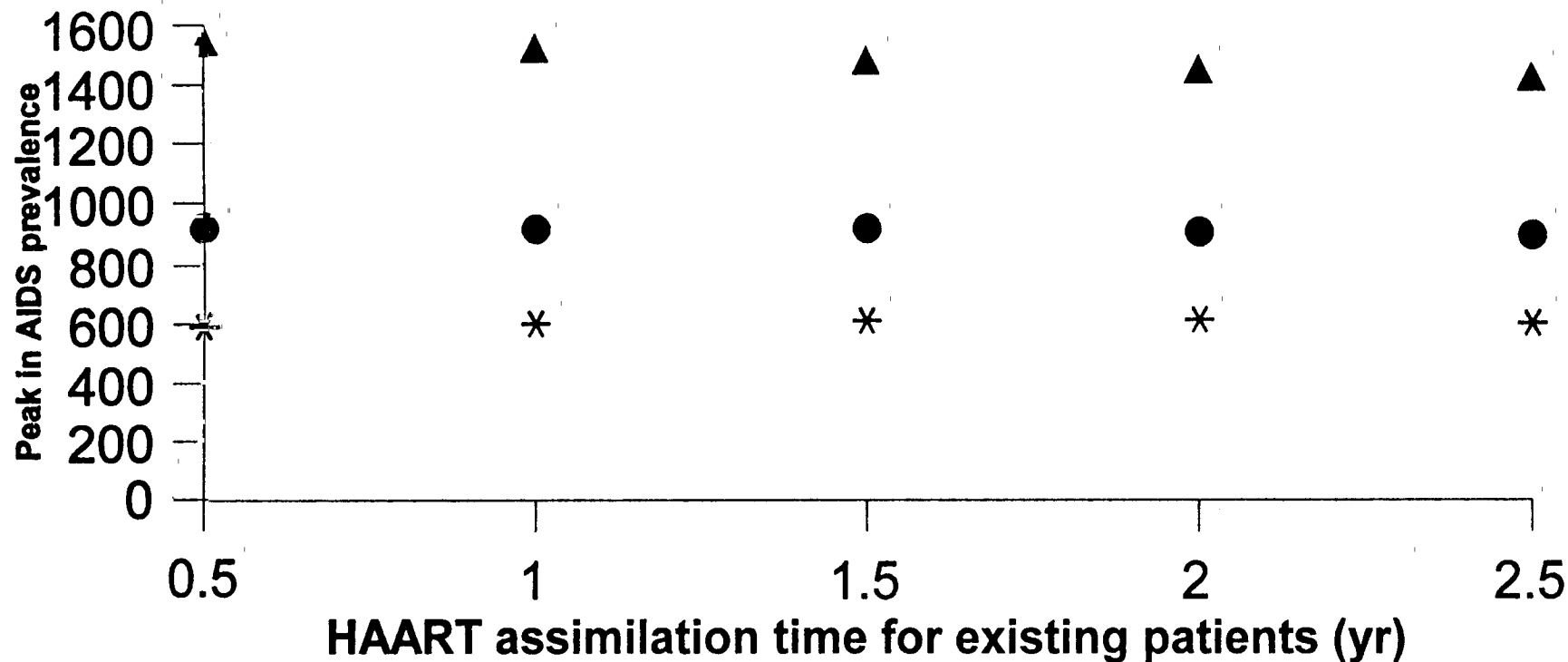
Figure 7: *Comparative scenarios of the effect on new AIDS cases of three assumed average efficacious periods for HAART*

Figure 8: *Comparative effects on the rebound peak in AIDS prevalence of changes in the assimilation time onto HAART and in the average efficacious period*

EPIDEMIOLOGICAL SCENARIOS FOR HAART IN THE U.K.
Assumed changes in number of partners by highly sexually active stratum

(a)

Baseline mean number of different partners p.a.

Approx 10% increase in mean number of different partners p.a.

Approx 30% increase in mean number of different partners p.a.

Approx 50% increase in mean number of different partners p.a.

EPIDEMIOLOGICAL SCENARIOS FOR HAART IN THE U.K.
Annual HIV incidence following a regression in behaviour

(b)

No initiation of HAART

Approx 10% regression in behaviour

Approx 30% regression in behaviour

Approx 50% regression in behaviour

Figure 9: *Assumed sustained regression in behaviour from 1997 by the most active sexual stratum (a); and the simulated results on new cases of HIV infection(b)*

horts in the early 1980's. Already research has been reported which suggests this is a phenomenon which could be evident in the homosexual risk group at large (Dilley, Woods and McFarland, 1997; Kelly et al, 1998; Van de Ven et al, 1999)

To investigate the consequences of such a regression in behaviour, the average number of partners per annum taken by the most sexually active stratum of homosexuals is assumed to increase, from 1997, by approximately 10%, 30% and 50%. By this is meant that the curve representing new partnership formations, which was assumed to be decaying very slowly following an extremely rapid phase of reduction after the health education campaigns in the 1980's and early 1990's, now turns upward. This takes the form of a smoothed exponential rise to a new equilibrium over a period of 4 to 5 years and the effect on prospective partnership formation can be seen in figure 9(a).

It should be remembered that this stratum represents the smallest stratum of susceptibles but certainly the largest stratum of infecteds. Prior to the introduction of HAART, this stratum will have been depleted through earlier mortality; it represented two-thirds of diagnosed AIDS cases by the mid-1990's. Also, at that time it accounted for an estimated 10,200 HIV sero-prevalent persons as compared with a total of 5,300 in the other two strata combined.

There is, therefore, plenty of scope for a serious regression in behaviour to manifest itself quite significantly and the results are dramatic in terms of new HIV infections (figure 9(b)). While the 10% change has relatively little effect, the 30% and, particularly, the 50% changes show a much more than proportionate effect on new HIV infections. If the regression is severe enough then the conditions are such that a new spiral in the epidemic is initiated. New infections spread rapidly amongst those recently HIV-negative (recall that the first stage of HIV pathogenesis has been shown to be the stage when infectivity is highest) and of course these new cohorts of infecteds will not be in receipt of any sort of viral suppression treatment unless and until they make themselves available for testing at a later point.

The gain in the positive loop has to be > 1.0 and then a new take-off is certain. If the regression in behaviour can be contained below a certain point then the effects on new infections (and the consequent costs of therapy if these new cases come forward for testing) will be mitigated.

Conclusions

Qualitative conclusions in respect of the operation and policy management of a dynamic system are the characteristic outcome of any system dynamics study. Even with the imperfections and uncertainties evident in the modelling work reported, it is possible here to derive a number of qualitative conclusions concerning issues surrounding the introduction of HAART as an HIV/AIDS therapy.

- New HIV infections will be considerably reduced deriving from the apparent very

low infectivity of those on HAART together with the increasing proportion of cases benefitting from the therapy. (Physicians are becoming more adept at handling individual tolerances.) Furthermore, this reduction in new infections will be unaffected by any eventual breakdown in efficaciousness of HAART especially if then the patients' physical condition most likely deteriorates very rapidly.

- The above conclusion will be compromised should it be shown that infectivity is only slightly increased over and above what is thought to be the case with patients on HAART. New infections are far more sensitive to changes in the infectivity probability than they are in changes to the duration of the efficacious period. Remember that HAART can transform the well-being of HIV/AIDS patients and so sexual relations are most likely to be resumed, with the associated risk of unprotected sex.

- A regression in sexual behaviour (in terms of the number of partners together with the risk of unsafe sex) will dent the benefits of HAART in terms of new HIV infections. This will involve a more than proportionate effect on HIV incidence the higher and more sustained is the regression in behaviour; it is akin to a new epidemic. In this scenario there will emerge a need for further budget allocations to handle the soaring drugs bill.

- An unmistakeable conclusion is that, although the duration of time for which HAART proves efficacious might be thought to have a considerable impact, it is far less important than infectivity and behavioural influences on the progress of the epidemic. There should be no underfunding of health education just because the new therapy is apparently successful.

- Because there was a fairly rapid assimilation of patients onto HAART in 1995/96 in both the UK and other industrialised nations, any eventual breakdown in the new therapy will most likely lead to a sudden surge in new cases of advanced HIV disease or AIDS. This will have repercussions for the hospitals (where many AIDS wards have been closed following the success of HAART) and the community of health care workers generally.

It is undeniable that the advent of triple (and even quadruple) antiretroviral therapy has had an enormous impact on the current health and prognosis for those with HIV disease. But as things stand this is a phenomenon evident solely in those industrialised nations which can afford the drugs. While the research reported above offers some food for thought for policy makers in the West, it should not be forgotten that most, if not all, developing nations are denied, due to cost, the possibility of reflecting on the consequences of any new therapy (Hogg *et al*, 1998). Their immediate task is to contemplate the effects on their economies, their health care systems and their sick and orphaned children.

Acknowledgements

References

[1] Aboulker JP and Swart AM. (1993). Preliminary Analysis of the Concorde trial. *Lancet* **341**: 889-890.

[2] Andreoni M, Sarmati L, Ercoli L, Nicastri E, Giannini G, Galluzzo C, Pirillo MF and Vella S. (1997). Correlation between changes in plasma HIV RNA levels and in plasma infectivity in response to antiretrovirals. *AIDS Research and Human Retroviruses* **13:7** 555-561.

[3] Beck EJ *et al.* (1999) Decreased morbidity and use of hospital services in English HIV-infected individuals with increased uptake of antiretroviral therapy 1996-97. *AIDS* **13** 2157-2164.

[4] Blakeslee D. (1999). A Boost for HAART. In: *Program and abstracts of the 6th Conference on Retroviruses and Opportunistic Infections* January 31 - February 4, Chicago, Ill.

[5] Carpenter CCJ, Fischl MA, Hammer SM, *et al.* (1996). Antiretroviral therapy for HIV infection in 1996: recommendations of an international panel. *Journal of the American Medical Association* **276**: 146-154.

[6] Carpenter CCJ (1997). Antiretroviral therapy for HIV infection in adults in 1997. 4th *Int. Congr. AIDS in Asia and the Pacific*, Manila.

[7] Carpenter CCJ, Fischl MA, Hammer SM, *et al.* (1997). Antiretroviral therapy for HIV infection in 1997: updated recommendations of the International AIDS Society USA panel. *Journal of the American Medical Association* **277**: 1962-1969.

[8] Carpenter CCJ, Fischl MA, Hammer SM, *et al.* (1998). Antiretroviral Therapy for HIV Infection in 1998: updated recommendations of the International AIDS Society USA Panel. *Journal of the American Medical Association* **280**: 78-86.

[9] Cohen MS. (1998). Sexual transmission of HIV-1: where epidemiology meets biology. 5th *Conference on Retroviruses and Opportunistic Infections*. Chicago, Ill. Abstract S34.

[10] Coombs R. (1998). HIV-1 compartmentalisation in the genital tract: implications for sexual transmission. 5th *Conference on Retroviruses and Opportunistic Infections*. Chicago, Ill. Abstract S33.

[11] Dangerfield BC and Roberts CA (1994). Fitting a Model of the Spread of AIDS to Data from five European Countries. In: Dangerfield BC and Roberts CA (eds). *O.R. Work in HIV/AIDS* (2nd edition). Operational Research Society: Birmingham. pp 7-13.

[12] Dangerfield BC and Roberts CA (1996a). An Overview of Strategy and Tactics in System Dynamics Optimisation. *Journal of the Operational Research Society* **47** 405-423.

[13] Dangerfield BC and Roberts CA (1996b). Relating a Transmission Model of AIDS Spread to Data: some International Comparisons. In: Isham V and Medley G (eds). *Models for Infectious Human Diseases: Their Structure and Relation to Data.* CUP: Cambridge. pp 473-476.

[14] Delmas MC, Schwoebel V, Heisterkamp SH, *et al.* (1995). Recent trends in *pneumocystis carinii* pneumonia as an AIDS-defining disease in nine European countries. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **9:1** 74-80.

[15] Dilley JW, WJ Woods and W McFarland (1997) Are Advances in Treatment Changing Views about High-Risk Sex? *New England Journal of Medicine* **337:7** 501-502. (Letter)

[16] Downs AM (1999) EuroHIV: Institut de Veille Sanitaire, Saint-Maurice, France (Personal communication).

[17] Ferguson NM, deWolf F, Ghani AC, Fraser C, Donnelly CA, Reiss P, Lange JMA, Danner SA, Garnett GP, Goudsmit J and Anderson RM. (1999). Antigen-driven CD4+ T cell and HIV-1 dynamics: residual viral replication under highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* **96:26** 15167-15172.

[18] Gulick RM, JW Mellors, D Havlir *et al.* (2000) 3-year suppression of HIV viræmia with indinavir, zidovudine and lamivudine. *Annals of Internal Medicine*, **133:1** 35-39.

[19] Heisterkamp SH, JC Jager, EJ Ruitenberg, JAM van Druten and AM Downs (1989). Correcting Reported AIDS Incidence: a statistical approach. *Statistics in Medicine* **8** 963-976.

[20] Hogg RS, Weber AE, Craib KJP, Anis AH, *et al.* (1998). One World, One Hope: the cost of providing antiretroviral therapy to all nations. AIDS **12:16** 2203-2209.

[21] Hoover DR (1995). The effects of long term zidovudine therapy and *pneumocystis carinii* prophylaxis on HIV disease. *Drugs* **49:4** 20-36.

[22] Kelly JA *et al* (1998) Protease inhibitor combination therapies and perceptions of gay men regarding AIDS severity and the need to maintain safer sex. *AIDS*, **12** F91-F95.

[23] Lewin S, Zhang L, Kostrikis R *et al.* (1999). The impact of HIV infection on the number of recent thymic emigrants in blood and the effect of HAART. In: *Program and abstracts of the 6th Conference on Retroviruses and Opportunistic Infections* January 31 - February 4, Chicago, Ill.

[24] Markson L, Cunningham W, Andersen R, Crystal S, Fleishman J, Fischl M, Gifford A, Nelson C, Liu H, Bozzette S, Shapiro M and Wenger N (1999). Who Receives Highly Active AntiRetroviral Therapy (HAART)? – Data from a nationally representative sample. In: *Program and abstracts of the 6th Conference on Retroviruses and Opportunistic Infections* January 31 - February 4, Chicago, Ill. Abstract 105.

[25] Perelson AS, Neumann AU, Markowitz M, Leonard JM, and Ho DD. (1996). HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**: 1582-1586.

[26] Perelson AS, Essunger P, Cao YZ, *et al.* (1997). Decay characteristics of HIV-1 infected compartments during combination therapy. *Nature* **387**: 188-191.

[27] Roberts CA and BC Dangerfield (1990). Modelling the Epidemiological Consequences of HIV Infection and AIDS: a contribution from Operational Research. *J Opl Res Soc* **41:** 273-289.

[28] Roberts CA and BC Dangerfield (1992) Estimating the Parameters of an AIDS Spread Model using Optimisation Software: results for two countries compared. In *System Dynamics 1992*, JAM Vennix, J Faber, WJ Scheper and CA Th. Takkenberg Eds., System Dynamics Society, Cambridge, Mass. 605-617.

[29] Sterman JD (2000) *Business Dynamics*, Irwin McGraw-Hill, Boston.

[30] Van de Ven P, S Kippax *et al* (1999). HIV Treatments, Optimism and sexual behaviour amongst gay men in Sydney and Melbourne. *AIDS* **13** 2289-2294.

[31] Vernazza PL, Gilliam BL, Dyer J, Fiscus SA, Eron JJ, Frank AC and Cohen MS. (1997). Quantification of HIV in semen: correlation with antiviral treatment and immune status. *AIDS* **11:8** 987-993.

[32] Wolinsky S. (1998). Virologic correlates of HIV-1 transmission. 5th *Conference on Retroviruses and Opportunistic Infections*. Chicago, Ill. Abstract S36.

[33] Wong JK, Hezareh M, Gunthard H *et al.* (1997). Recovery of replication-competent HIV despite prolonged suppression of plasma viræmia. *Science* **278**: 1291-1295.