A Dynamic Model of Drug Initiation: Implications for Treatment and Drug Control [+]

Doris A. Behrens [*]
Vienna University of Technology
Department of Operations Research and Systems Theory

Jonathan P. Caulkins [°]
Carnegie Mellon University
H. John Heinz III School of Public Policy and Management
and
RAND Drug Policy Research Center

Gernot Tragler, Josef L. Haunschmied, and Gustav Feichtinger
Vienna University of Technology
Department of Operations Research and Systems Theory

Abstract:
We set up a time-continuous version of the first-order difference equation model of cocaine use introduced by Everingham and Rydell (1994) and extend it by making initiation an endogenous function of prevalence. This function reflects both the epidemic spread of drug use as users „infect" non-users and Musto's (1987) hypothesis that drug epidemics die out when a new generation is deterred from initiating drug use by observing the ill effects manifest among heavy users. Analyzing the model's dynamics suggests that drug prevention can temper drug prevalence and consumption, but that drug treatment's effectiveness depends critically on the stage in the epidemic in which it is employed. Reducing the number of heavy users in the early stages of an epidemic can be counter-productive if it masks the risks of drug use and, thereby, removes a disincentive to initiation. This strong dependence of an intervention's effectiveness on the state of the dynamic system illustrates the pitfalls of applying a static control policy in a dynamic context. Inasmuch as past research has concluded that treatment is a cost-effective intervention with no qualifications as to the timing of the intervention, it also illustrates the pitfalls of a static policy analysis of a dynamic problem.

Keywords: Nonlinear dynamic systems, Hopf bifurcations, illicit drugs

---

[+] This research was partly financed by the Austrian Science Foundation under contract No. P11711-SOZ, the National Consortium on Violence Research, and by the US National Science Foundation under Grant No. SBR-9357936. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
[*] Argentinierstraße 8/119, A-1040 Vienna, Austria; email: dbehrens@e119ws1.tuwien.ac.at
[°] 5000 Forbes Ave., Pittsburgh, PA 15213; email: caulkins+@andrew.cmu.edu

1 Introduction

Illicit drug use and related crime have imposed significant costs on the US and various source and transshipment countries for a number of years. More recently, drug problems have grown in other industrialized countries to the point that, in Stares' (1996) terms, drugs have become a „global habit." Hence it is important to understand drug use and how it responds to drug control interventions. This paper contributes to that effort by introducing a continuous time model of drug demand that incorporates a feedback effect of the current prevalence, or level, of use on initiation into new use. Analyzing the model generates important new insights for how epidemics of drug use should be studied and controlled. The model is parameterized with data from the US cocaine epidemic because of its magnitude and because data on that epidemic are relatively good, but it is at least plausible that the qualitative conclusions generalize to similar drugs in similar contexts.

Before describing the model, we provide a brief sketch of the US cocaine epidemic and the relevant literature we seek to extend. Cocaine use in the US started growing in the late 1960s. The number of users peaked in the early 1980s, but total consumption remained near its mid-1980s peak for more than a decade as the proportion of so-called „heavy" users grew. Thus trends in total consumption did not, and need not, mirror trends in overall prevalence. This raises the question of what summary measure best reflects trends in the magnitude of a drug problem.

Though prevalence is often used as such a summary measure, it is not a particularly good one (Caulkins and Reuter, 1997). There are many other possible measures, including how much users spend on drugs, the number of people requiring treatment, drug-related crime, drug arrests, drug-related social costs, and the number of drug-related emergency room episodes. Rydell et al. (1996) argue, though, that the quantity or weight consumed has advantages as a general purpose scalar measure of the size of a drug problem.

Since there is enormous heterogeneity across users in rates of consumption and the average rate of consumption over a population can change over time, tracking trends in total consumption requires modeling separately the numbers of users at different levels, or intensities, of drug use. Ideally one would model the whole spectrum of consumption behavior, from occasional use in small amounts up to frequent use in large amounts, but data limitations make that infeasible.

Everingham and Rydell (1994) recognize this tension and suggest that, at least for cocaine, a simple dichotomous distinction between „light" and „heavy" users is sufficient. They operationalize the distinction using data from the National Household Survey of Drug Abuse (NHSDA), which measures the prevalence of cocaine use among the US household population. In particular, people who report using cocaine „at least weekly" are defined to be heavy users, while those who consumed at least once within the last year but used less than weekly are called „light users." The average heavy user consumes cocaine at a rate approximately seven

times that of an average light user and exhibits substantially greater adverse consequences associated with that drug use.

Using this definition and data on drug use among the household, homeless, and incarcerated populations, Everingham and Rydell set up a Markovian model of population flows from non-use into light use and then into and out of light and heavy use. The transition parameters that determine the flows between those states were selected to match the historical data. They used this model to understand what has happened to date in the current cocaine epidemic, to project the future under different initiation scenarios, and to compare effectiveness of drug treatment and three types of drug enforcement (Everingham and Rydell, 1994; Rydell and Everingham, 1994).

A significant limitation of the Everingham and Rydell model is that initiation is scripted. Future projections and policy simulation exercises are predicated on a fixed projection of future initiation that is insensitive to the course of the drug epidemic. That is problematic because initiation rates are significantly influenced by the current prevalence, or level, of use. In particular, most people who start using drugs do so through contact with a friend or sibling who is already using. Indeed, the metaphor of a drug „epidemic“ is commonly used precisely because of this tendency for current users to „recruit“ new users. If that were the only mechanism by which current use affected initiation one might expect initiation to increase monotonically. Musto (1987) has argued that, in addition, knowledge of the possible adverse effects of drug use acts as a deterrent or brake on initiation. He hypothesizes that drug epidemics eventually die out when a new generation of potential users becomes aware of the dangers of drug abuse and, as a result, does not start to use drugs. Whereas many light users work, uphold family responsibilities, and generally do not manifest obvious adverse effects of drug use, a significant fraction of heavy users are visible reminders of the dangers of using addictive substances. Hence, one might expect large numbers of heavy users to suppress rates of initiation into drug use.
Thus the existence of large numbers of heavy users is in some sense not only bad, because they consume at high rates and impose large costs on society, but also good because they tend to discourage initiation. Heavy users impose costs in the near term, but they generate a perverse sort of „benefit“ for the future by reducing current initiation and thus future use. This generates a tension or interplay between treatment programs that reduce the number of heavy users and prevention programs that try to discourage new (light) users. Furthermore, the impact of any given heavy user on initiation likely depends on the number of other heavy and light users, so the magnitude of the „benefit“ of a heavy user depends on the stage of the drug epidemic. So, logically, one would expect the effectiveness of treatment, and perhaps prevention, to depend on the stage of the epidemic.

Omitting these feedback effects of current prevalence on initiation is of relatively little consequence if the goal is to analyze the effectiveness of treatment and enforcement at a particular point in time, as Rydell and Everingham did. It is of

enormous consequence, however, for understanding how effective prevention programs are or for understanding how the effectiveness of an intervention such as treatment might vary over the course of an epidemic. Hence, these issues are the focus of our analysis.

2 A model of the feedback effect of prevalence on initiation

From the above it is clear that the initiation rate should be increasing in the number of light users and decreasing in the number of heavy users, but there are many such functions. It seems plausible that a reasonable model might have the following additional properties.

1) The rate at which current users „recruit“ initiates is proportional to the number of light users. It is assumed that heavy users do not recruit initiates because they manifest ill effects of drug use and/or because they have been using long enough that they are older and socially distant from youth in the prime initiation ages.

2) The rate at which current light users „recruit“ initiates is moderated by the „reputation“ or image the drug has, and the reputation is governed by the relative number of heavy and light users not the absolute number of heavy users. Even if there were a number of heavy users, the drug might appear benign if they were buried in a sea of (relatively happy) light users.

3) Although most new users are „recruited,“ for others the impetus to use is internal. In the jargon of diffusion models, these individuals are „innovators“ who initiate on their own for the sake of curiosity, by shifting from other drugs, or for some other reason, but not through the urging of someone who is already a user.

Peter Rydell (personal communication) suggested the following endogenous initiation function as one that has these characteristic features. (Since all variables depend on the current stage of the epidemic, we omit an explicit denotation of the dependence on time.)

$$I(L,H) = \mathrm{t} + sL\exp\left[-q\frac{H}{L}\right] \qquad (1)$$

$L>0$ ... number of light users
$H>0$ ... number of heavy users
$s=0.61$ ... average rate at which light users attract non-users
$q=7.0$ ... constant measuring the deterrent effect of heavy drug abuse
$\mathrm{t}=50{,}000$ ... number of innovators

The values of parameters *s* and *q* were found by minimizing the sum of the squared difference between the modeled and observed annual cocaine initiation rates from 1970 to 1991.[1] Reasonably good fits can be obtained for a wide range of values. Roughly speaking, t should be approximately *50,000*, *s* and *q* should be chosen such that they are located in the „white region" of Figure 1, where larger values of *s* are associated with larger values of *q*. Our base case estimates are just one plausible set of initiation parameters, so we conduct sensitivity analysis with respect to those values below.

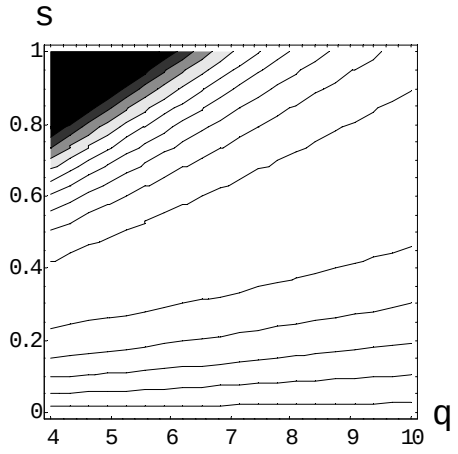

Figure 1

Contour plot for parameters *s* and *q* depicting isoclines for the sum of the squared differences between modeled initiation and observed data

The size of parameter *s=0.61* can be interpreted as indicating that approximately two light users would „persuade" one non-user per year to try cocaine, if there were no heavy users giving cocaine a bad reputation. The constant *q* corresponds to the ratio of the average heavy user's consumption rate to that of the average light user.

As Figure 2 shows, the modeled and historical initiation curves are reasonably close. This of course by no means validates this particular initiation function. Though the observed data do not reject the assumed functional form based on regression analysis of the transformed series[2], other functional forms with about the same number of parameters and the same or similarly appealing properties produce comparably close fits to the historical data. (Caulkins et al., 1997.) True validation is very difficult, however, given the inability to collect further data.

[1] Following Everingham and Rydell (1994, pp.31-33), we estimate historical initiation rates by taking the average of so-called „difference" and „retrospective" estimates (Gfroerer and Brodsky, 1992) but substituted the newer retrospective estimates provided by Johnson et al. (1996) for those used by Everingham and Rydell. Estimates of rates of light and heavy use are taken from Everingham and Rydell (1994).

[2] We checked the significance of the regression parameters with the usual F-test. The corresponding p-value is 10^{-6}.

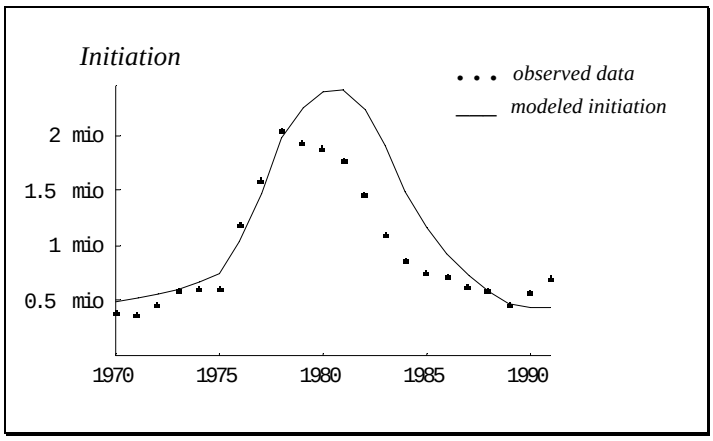

Figure 2

Plot of observed and modeled initiation for parameter-set:
s=0.61, q=7.0, t *=50,000*

The rest of our model is essentially a continuous time analogue of Everingham and Rydell's model. Due to the heterogeneity in cocaine consumption behavior, Everingham and Rydell divided the US population into three groups: non-users, light users and heavy users. Because the number of non-users is very large compared to the number of users it behaves like a constant and need not be modeled explicitly (Everingham, Rydell and Caulkins, 1995). (See Figure 3.)

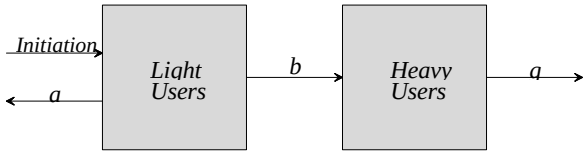

Figure 3

Flow diagram for model *(2)*

The flow rates from one state to another are assumed to be proportional to the source states and are computed as the time-continuous equivalents of the Everingham-Rydell estimates (1994, p.43). For example, they estimate that 15% of light users quit use each year and 2.4% escalate to heavy use, so we set $a = -\ln[1 - 0.15] @0.163$ and $b = -\ln[1 - 0.024] @0.024$, to three decimal places.

The one difference between this model and that of Everingham and Rydell is that Everingham and Rydell divided the outflow from heavy use into a flow out of use altogether (*g*) and a flow back into light use (labeled *f*). We drop the latter flow for both theoretical and practical considerations. Theoretically, a flow from heavy to light use coupled with the Markov assumption implies that former heavy users who have de-escalated to light use and light users who had never been heavy users are indistinguishable. But probably it is easier to relapse to heavy use than to enter the

state for the first time. Hence we prefer to have only a flow from heavy use to non-use and view that rate as net of relapse.

Practically, Everingham and Rydell found that the data did not identify *f* and *g* individually. Any combinations of these parameters such that $f + g = 0.06$ fit the historical data reasonably well. Since reducing the number of flows simplifies the analysis, we choose $f = 0$ and $g = -\ln[1 - 0.06] = 0.062$.

Now we can define a time-continuous dynamical Everingham & Rydell-type model for the evolution of cocaine consumption in terms of prevalence (number of light and heavy users) including the feedback effect of prevalence on initiation which is described by function *(1)*:

$$\begin{aligned} \dot{L} &= I(L,H) - (a+b)L, & L(0) &= L_0 \\ \dot{H} &= bL - gH, & H(0) &= H_0 \end{aligned} \qquad (2)$$

L>0 ..... number of light users
H>0 ..... number of heavy users
I(L,H) ..... initiation into light use
a@0.163 ..... average rate at which light users quit
b@0.024 ..... average rate at which light users escalate to heavy use
g@0.062 ..... average rate at which heavy users quit

3 Analysis of the Model

3.1 Effects of Changes in the Flow Rates on the Equilibrium

In order to study the global qualitative properties of the solutions of system *(2)*, we construct the phase diagram (Figure 4) for the base case set of parameters $a \cong 0.163,\ b \cong 0.024,\ g \cong 0.062,\ s = 0.610,\ q = 7.0,\ t = 50{,}000$. The trajectories spiral counter-clockwise into a focus. This behavior is not obvious from Figure 4 because the trajectories decay so quickly toward the equilibrium, but it is easy to see in an expanded view of the region surrounding the equilibrium (not shown).

The locus of points where the number of heavy users remains constant is characterized by the constant ratio $H/L = 0.39$ and is depicted as the straight line $H = 0.39L$ in Figure 4.[3] For $H/L > 0.39$ the number of heavy users decreases, and for $H/L < 0.39$ it increases. The curve which defines the locus of points where the number of light users remains unchanged is characterized by $H = -0.143L \ln\left[0.31 - \frac{81{,}967.2}{L}\right]$ and asymptotically approaches $L = 267{,}380$. Below and to the left of this $\dot{L} = 0$ isocline light use

[3] For a general derivation of the formulas see Appendix A.1.

tends to increase, and above this curve light use decreases. All trajectories starting in the positive quadrant spiral towards the unique equilibrium

$$\hat{E} = \begin{pmatrix} \hat{L} \\ \hat{H} \end{pmatrix} = \begin{pmatrix} 338{,}416 \\ 132{,}864 \end{pmatrix}.$$

This characterization of $\hat{E}$ as a stable focus can be confirmed locally by linearization of the differential equations *(2)* at the equilibrium. This yields a pair of conjugate complex eigenvalues with negative real parts, $l_{1,2} \cong -0.051 \pm 0.081i$. The amplitude is determined by the size of the real parts of the characteristic roots, and the frequency by the magnitude of the imaginary parts.

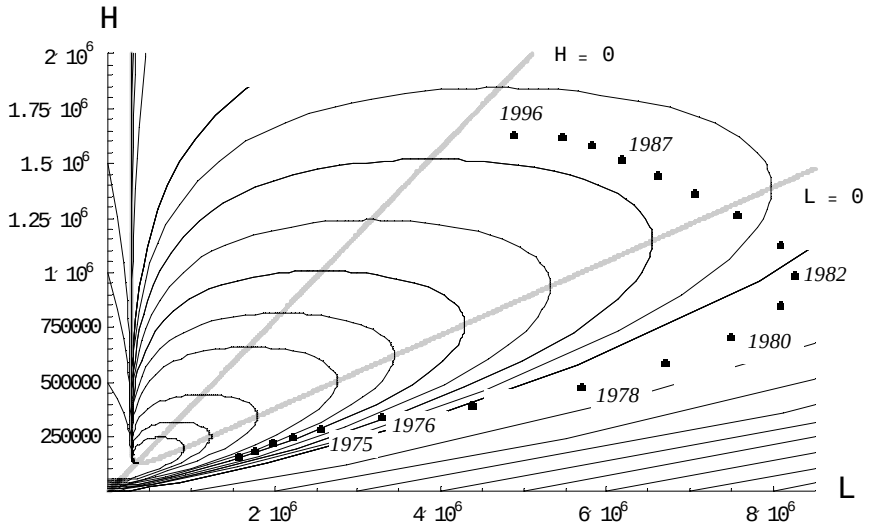

Figure 4

Phase portrait of model *(2)* for the base case parameter-set:
$a \cong 0.163$, $b \cong 0.024$, $g \cong 0.062$, $s = 0.61$, $q = 7.0$, $t = 50{,}000$, including smoothed historical trajectory of the current US cocaine epidemic (Everingham&Rydell)

In equilibrium, the proportion of users who are heavy users is $b/(b+g) = 0.28$. That density of heavy users is sufficient to deter over 93% of potential recruited initiation. Just to avoid any misunderstanding, we do not claim that having so many heavy users – and the associated problems – is good. But their deterrent effect on initiation might help enhance or perhaps even replace preventive programs at certain stages of the epidemic.

Figure 4 also shows the historical trajectory of the cocaine epidemic which to date has followed a spiral not so different from that which the model predicts. The modeled and historical data also move around their respective paths at nearly the same rate. The fit is not perfect; the historical data reflect a higher, sharper peak in light use. Nevertheless, the similarity is striking given that the actual epidemic has been subject to a varying set of drug control interventions over time that could be

responsible for deviations from the model's uncontrolled path. Likewise, idiosyncratic historical events, such as Len Bias' death and the sharp increases in prices in late 1989, could account for some of the difference between historical and modeled data.

3.2 Sensitivity Analysis

A range of parameter values fit the data reasonably well, so we are interested in understanding how variations in the parameters around their base case values affect the system, both with respect to the general system behavior and also with respect to levels of drug consumption. With respect to system behavior, we want to know what kind of perturbations in the parameters are necessary to alter the damped oscillatory behavior depicted in Figure 4. Structurally we observe four behaviors: damped oscillation, exponential decay toward the equilibrium, stable limit cycles, and exponential growth. The last is not interesting because it does not match historical data and cannot continue indefinitely in a finite population.

Stable limit cycles may emerge in a supercritical Hopf bifurcation if any of the parameters - except for the number of innovators, t - is deviated sufficiently from its base case value.[4] The Hopf theorem (see Appendix A.2) states that limit cycles exist if (i) two purely imaginary eigenvalues exist for a critical value of the parameter, such that (ii) the imaginary axis is crossed with nonzero velocity. For *a* and *s* the Hopf bifurcation values could be calculated explicitly. For parameters *q*, *b* and *g* the conditions (Equations (A5) and (A6)) are given implicitly and must be calculated numerically.

Table 1: Effect of Parameter Values on System Behavior

Parameter	t	*a*	*s*	*q*	*b*	*g*
Base case value	50,000	0.163	0.610	7.000	0.024	0.062
Everingham & Rydell Range	NA	0.157 - 0.168	NA	NA	0.020 - 0.030	0.000 - 0.094
Range yielding exponential decay	None	0.348 - 1.000	0.000 - 0.166	>11.440	0.038 - 1.000	0.000 - 0.043
Range yielding damped oscillations	All	0.060 - 0.348	0.166 - 1.036	5.080 - 11.440	0.018 - 0.038	0.043 - 0.094
Range yielding stable limit cycles	None	0.021 - 0.060	1.036 - 1.631	3.679 - 5.080	0.013 - 0.018	0.094 - 0.138
Range yielding exponential growth	None	0.000 - 0.021	> 1.631	3.014 - 3.679	0.120 - 0.013	0.138 - 0.143
Hopf critical value	None	0.060	1.036	5.080	0.018	0.094

[4] Note that we are able to exclude the case of subcritical Hopf bifurcations for each of the system-parameters. (see Appendix A.3).

Table 1 shows the ranges of the parameters for which Everingham and Rydell (1994, p.42) found good fits to the historical data and the ranges yielding each of the system behaviors. In general, the plausible parameter ranges are comfortably within the ranges which yield damped oscillatory behavior. The exception is parameter *g*. Section 3.3 discusses what happens if *g* begins to exceed its critical value.

Table 2: Effect of Parameter Values on Consumption

Parameter	t	*a*	*s*	*q*	*b*	*g*
Elasticity of $Q_{50}(1967)$	0.006	- 0.051	0.101	- 0.064	- 0.053	0.011
Elasticity of $Q_{50}(1978)$	0.001	- 0.038	0.066	- 0.048	- 0.031	0.006
Elasticity of $Q_{50}(1982)$	$8\text{x}10^{-4}$	- 0.020	0.026	- 0.025	- 0.090	$3\text{x}10^{-4}$
Elasticity of $Q_{50}(1985)$	$6\text{x}10^{-4}$	- 0.007	0.004	- 0.007	0.001	- 0.004
Elasticity of $Q_{50}(1996)$	$6\text{x}10^{-4}$	- 0.003	$4\text{x}10^{-4}$	- 0.001	0.002	- 0.005
Elasticity of equilibrium consumption	0.010	- 0.011	0.003	- 0.007	- 0.004	$3\text{x}10^{-5}$

Concerning the sensitivity of consumption to parameter values, it is useful to consider both consumption in steady state and discounted total consumption over a relevant finite planning horizon such as 50 years, denoted by $Q_{50}(year)$[5], starting with initial data characteristic of the US cocaine epidemic in that year. (Following Everingham and Rydell, we assume light and heavy users consume at an annual rate of 16.42 grams and 118.93 grams per year, respectively, and discount at 4% per annum.) Table 2 reports the elasticity of those quantities with respect to each of the six parameters, where consumption appears less sensitive to parameter perturbations at the latter stages of the epidemic. The implications are as follows.

Parameters Governing Initiation

Although the number of innovators, t, has no effect on the stability behavior of the system whatsoever, the size of the equilibrium state is proportional to the number of innovators. (See Equation (A3) in the appendix.)

The effects of changes in the rate at which light users attract non-users, *s*, can also be derived from the Equations (A3) and (A4). The equilibrium number of users declines with declining *s*. All values between *s*@0.166 and *s*@1.036 lead to damped oscillations, but the smaller *s* is, the sooner the epidemic peaks and the less intense the epidemic will be. This shows the importance of teaching non-users to resist peer pressure to use drugs. Especially at early stages of the epidemic cutbacks in *s*,

[5] $Q_{50}(year) = \int_0^{50} e^{-0.04t}\left(16.42L(t) + 118.93H(t)\right) dt$ subject to model *(2)*, provided with initial data from the cocaine epidemic observed for the corresponding year.

as well as increments in *a*, which are both goals of prevention programs, can substantially reduce the magnitude of the cocaine epidemic.

Increasing *q*, which measures the deterrent effect of heavy users on initiation, has the expected effect of diminishing both the steady state level and the intensity of the epidemic. Since the values of *q* and *s* are linked (large values of *q* correspond to large values of *s*), large one-sided deviations from the base case parameter values are not considered.

Parameters Pertaining to Light Use

From Equation (A3) we can derive that a reduction in the rate at which light users quit, *a*, raises the steady state values, and according to Equation (A4) the amplitude and duration of the epidemic grow too. Particularly in the beginning of the epidemic consumption is quite sensitive to changes in *a*. Even though the Everingham and Rydell interval is narrow (*0.157-0.168*), deviations within that range lead to deviations in total consumption over 50 years (starting with data from 1967) of up to approximately 21.5% from their base values. In contrast, steady state values change no more than 3.9%.

Parameters Pertaining to Heavy Use

For initial data from the beginning of the epidemic the model's behavior is very sensitive to changes in *b* and *g*, the two flow rates affecting heavy use. With respect to parameter *b*, reducing escalation to heavy use would at first seem to be good because heavy users consume at much higher rates than light users and persist longer in that use. On the other hand the fewer light users who escalate to heavy use, the more light users there are to attract non-users into use and the fewer heavy users there are to deter initiation.

If light users progress to heavy use rapidly (more than 3.7% of them per year), then the system approaches the steady state directly, without overshoot or oscillation. If very few light users progress to heavy use each year (less than 1.8%), then the system would not approach a single positive equilibrium. However, such values are outside the range Everingham and Rydell found to fit the historical data.

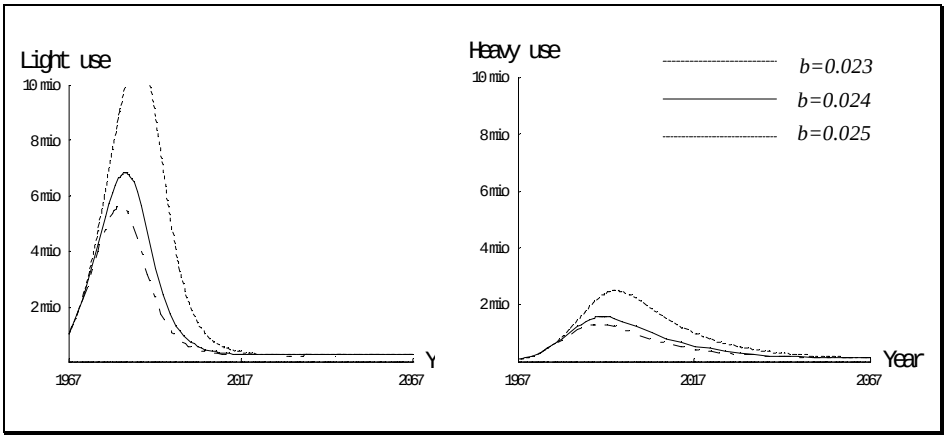

Figure 5

Time paths of light and heavy users for different values of the rate at which light users escalate to heavy use, *b*, for initial data from 1967

Even though variations in *b* do not threaten the basic system behavior, they do dramatically affect consumption for initial data from the beginning of the current cocaine epidemic. Figure 5 shows that small changes in the value of parameter *b* lead to large changes in the height of the peaks. If we increase *b* by *0.001* (an increase of 4.2%, depicted by the dashed line in Figure 5), $Q_{50}(1967)$ is reduced by 14%. The corresponding rise in the steady state number of heavy users is 0.4%. This increases deterrence, reducing the equilibrium number of light users by 2.5%. On the other hand, if we decrease *b* by *0.001* (which is depicted by the dotted line in Figure 5) the equilibrium number of light users will increase by 5.3% while the equilibrium number of heavy users decreases by 0.3%. More dramatically, it leads to a 45% increase in total consumption over a 50 year time horizon because it tremendously increases the intensity of the initial epidemic.

Increasing the rate at which heavy users quit at early stages of the epidemic has the opposite effect that one would at first expect. Increasing *g* throughout the entire epidemic, which corresponds to treating a larger proportion of heavy users at all times, actually increases both the steady state number of heavy users and, even more dramatically, consumption over the first 50 years, for initial data observed in 1967. For example, if we increase *g* from *0.062* to *0.09*, there will be 56% more light users and about 7% more heavy users in steady state, and consumption over the first 50 years will be 121% higher. On the other hand, reducing the rate at which heavy users quit reduces consumption in the long run by increasing the number of heavy users who act as deterrent to initiation. E.g. decreasing *g* from *0.062* to *0.043* (dotted line in Figure 6) reduces the number of light users in steady state by 15.6%, increases the number of heavy users by 21.4%, and cuts consumption over the first 50 years by 20.5%.

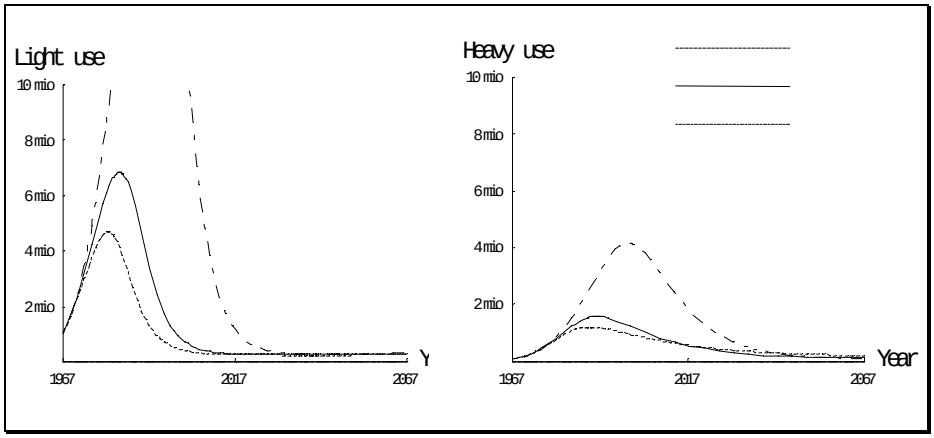

Figure 6

Time paths of light and heavy users for different values of the rate at which heavy users quit, *g*, for initial data from 1967

This by no means implies that treatment is never useful or that it has no role to play in drug control. Treatment at latter stages of the epidemic can help reduce drug consumption. The seemingly perverse result applies when one is obligated to pursue a given level of treatment throughout the entire course of the epidemic. Treating and removing heavy users when there are relatively few heavy users is counter-productive because each user contributes significantly to deterring initiation. If there are many heavy users relative to the number of light users, however, treating and removing some does not greatly jeopardize the negative reputation of the drug or the suppression of initiation.

Figure 7 makes this clear. It distinguishes combinations of numbers of light and heavy users for which additional treatment beyond the level historically pursued in the US (that which is reflected in the flow parameters) is counter-productive and combinations for which it is helpful. Within the region where additional treatment is helpful, the figure shows isoclines indicating combinations for which expanding treatment reduces consumption by a particular amount Δ per additional person treated. (For the calculation of Δ see Appendix A.4)

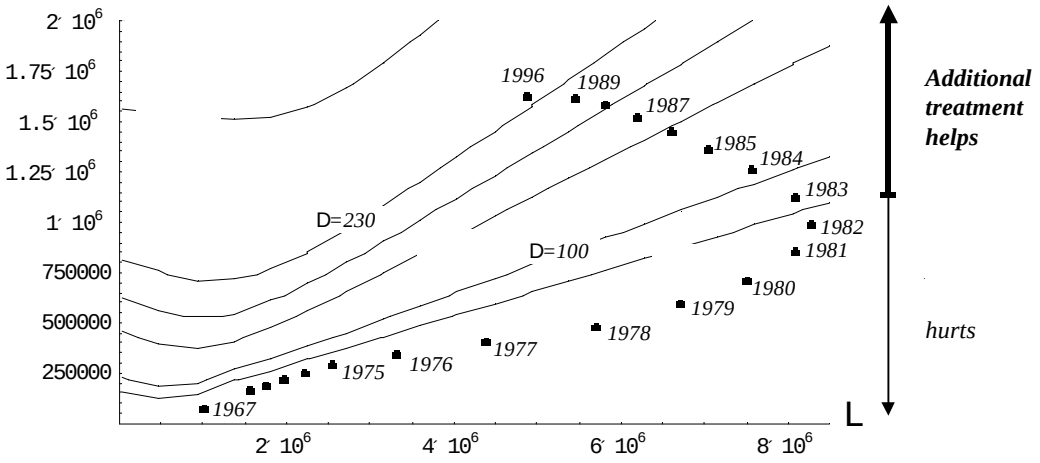

Figure 7

Contour plot for light and heavy users depicting isoclines for which additional treatment ($u = 31\% \rightarrow \tilde{u} = 39.2\%$ which implies $g = 6.2\% \rightarrow \tilde{g} = 7.2\%$) reduces consumption by Δ grams per additional heavy user treated; including smoothed historical trajectory of the current US cocaine epidemic (Everingham&Rydell)

The figure can be used as a guide to policy makers. If the epidemic is currently within the region where more treatment is counter-productive, then treatment levels should be below those historically observed in the US. If the epidemic is in the region where further treatment is effective, then the policy maker would have to estimate the cost of treating a user (perhaps around $2,000) and judge the benefit of reducing cocaine consumption by a certain amount. (The reduction in social cost per gram of consumption averted might be on the order of $100.) Together these

figures determine how many grams of consumption a treatment must avert in order for additional treatment to be cost-effective. If the epidemic is above the corresponding isocline in Figure 7, then additional treatment is cost-effective.

3.3 Periodic prevalence

As mentioned above, for values of the rate at which heavy users quit in the interval $0.043<g<0.094047$ we always find damped oscillations, but if g reaches 0.094947 a stable limit cycle emerges. Although the base value (0.062) is below the critical value ($g_c=0.094047$), for two reasons it is plausible that the actual value of that flow parameter might approach or exceed the critical value.

First, some people think the Everingham and Rydell values of g are too low (more precisely their values of $f+g$ are too low). They imply an average subsequent career length for a heavy user of $1/(f+g) = 15$ years or so, which some view as implausibly long. Second, the outflow rate g is affected by the proportion of heavy users who receive treatment. Everingham and Rydell assumed that historically about 31% of heavy users have received treatment each year and that 13% of those treated leave heavy use. Hence, the outflow rate can be viewed as being $g = 0.062 + 0.13(u - 0.31)$, where u is the proportion of heavy users receiving treatment each year. If that proportion increased to 55.7%, g would reach 0.094047.

When only „few“ heavy users leave their source state ($g<g_c$) the deterrent effect of heavy drug abuse is sufficient to suppress initiation into light use and consequently escalation to heavy use over time. The epidemic declines. Contrary if a high percentage of heavy users receives treatment ($g>0.138$) and leaves the source state there is almost no deterrence at all and light users recruit non-users with rate s. Prevalence grows and grows. Between these values of g ($g_c<g<0.138$) however the deterrent force of heavy users is not strong enough to prevent further initiation, but it is strong enough to prevent nearly proportional initiation.

When $g>g_c$, the unique equilibrium (Equation (A3)) changes its stability because the real parts of the characteristic roots (Equation (A4)) become positive and (A6) is larger than zero, $D@1.24$, implying that the eigenvalues cross the imaginary axis with positive velocity. We obtain a supercritical Hopf bifurcation (see Appendix A.3), where orbits emerge for $g>g_c$. The fixed points $\hat{E}(g)$ have the local property of unstable foci and the periodic orbits are attracting.

Hence, a deviation beyond the limits of the Everingham-Rydell interval could result in periodic prevalence. Note that close to the bifurcation point the cycles „grow“ with the size of $\sqrt{g - 0.094047}$, until g reaches 0.138, which corresponds to the hypothetical case of 89.5% of heavy users receiving treatment each year. (For even larger proportions of heavy users receiving treatment the cycles disappear and the modeled system dynamics go off to infinity.)

As Figure 8 depicts, for a flow rate only infinitesimally larger than *0.094047* the numbers of light and heavy users would cycle counter-clockwise forever with a period of approximately 70 years (according to Equation (A9)). Since such cycling in a social system is usually undesirable, this is further evidence that treating a constant, large proportion of heavy users throughout an epidemic regardless of its course might not be the best policy.

Figure 8

Cyclical prevalence for parameter set $a=0.163$, $b=0.024$, $s=0.61$, $q=7.0$, $t=50{,}000$ and for different values of the rate at which heavy users quit, $g_C=0.094047<g<0.107$

4 Conclusion

Our analysis of a continuous analog to the Everingham and Rydell model of cocaine demand, augmented with an endogenous initiation function, generates a number of interesting observations.

Proposition 1
For the base case parameter values the model follows a curve similar to that observed in historical data and approaches a steady state which is relatively small in size but different from zero. The number of steady state users is proportional to the number of innovators, t, who initiate of their own accord, not at the urging of current users, although the way the steady state is reached does not depend on t.

Proposition 2
Damped oscillations always occur for flow rates within the limits given by Everingham and Rydell and for a broad range of initiation function parameters. The amplitude and frequency of the oscillations depend on the magnitude of the flow rates.

Proposition 3
In equilibrium about one-quarter of users are heavy users, and that proportion of heavy users deters most of the potential recruited initiation.

Proposition 4
Reducing the rate at which light users attract non-users, *s*, and/or increasing the rate at which light users quit, *a*, which both can be affected by prevention programs, can reduce the size of the modeled cocaine epidemic.

Proposition 5
Reducing the rate at which light users escalate to heavy use, *b*, and/or increasing the rate at which heavy users quit, *g*, increases the intensity of the modeled epidemic for initial data observed when the problem is arising.

Proposition 6
For values of the rate at which heavy users quit, *g*, just slightly beyond the limits of the Everingham-Rydell interval (specifically at g_c *@0.094047*) stable periodic solutions occur in a supercritical Hopf bifurcation. Deviations of this size are plausible and could be generated by treating large proportions of heavy users throughout the epidemic. Since such cycles are generally undesirable, this implies that treatment, though valuable at some times, should be targeted strategically to certain stages of a drug epidemic.

It is clear from these observations that the effectiveness of drug control interventions, notably treatment, can vary over the course of an epidemic. Different strategies are probably most effective at different specific stages of an epidemic, and one would expect the optimal mix of interventions to depend on the course and status of the epidemic. We hypothesize that prevention programs might be most effective in early stages of the epidemic, when most users are light users, while treatment programs may be most effective when a greater fraction of users are heavy users, as is typical in the latter stages of an epidemic.

Inasmuch as it makes sense to vary the mix of prevention and treatment over the course of an epidemic, investigations on intertemporal allocations to drug control programs would be an appropriate topic for further research using tools from the field of optimal control. Additionally, one could compare the influence of different drug control budget rules on intertemporal variation in these strategies.

References

Caulkins, J.P. and Reuter, P. (1997). „Setting Goals for Drug Policy: Harm Reduction or Use Reduction," *Addiction*, Vol. 92, No. 9, pp.1143-1150.

Caulkins, J.P., Rydell, C.P., Everingham, S.S., Chiesa, J., and Bushway, S. (1997) „An Ounce of Prevention a Pound of Uncertainty: The Cost-Effectiveness of School-Based Drug Prevention Programs." Mimeo.

Everingham, S.S. and Rydell, C.P. (1994). „Modeling the Demand for Cocaine", *MR-332-ONDCP/A/DPRC*, Santa Monica, CA: RAND.

Everingham, S.S., Rydell, C.P., and Caulkins, J.P. (1995). „Cocaine Consumption in the US: Estimating Past Trends and Future Scenarios," *Socio-Economic Planning Sciences*, Vol. 29, No. 4, pp.305-314.

Gfroerer, J. and Brodsky, M. (1992). „The Incidence of Illicit Drug Use in the United States, 1962-1989", *British Journal of Addiction*, No. 87, pp.1345-1351.

Guckenheimer, J. and Holmes, P. (1983). „*Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields"*, New York, NY: Springer.

Johnson, R.A., Gerstein, D.R., Ghadialy, P., Choy, W., and Gfroerer, J. (1996). „Trends in the Incidence of Drug Use in the United States, 1919-1992", Substance Abuse and Mental Health Services Administration, Office of Applied Studies, US Department of Health and Human Services.

Musto, D.F. (1987). „*The American Disease: Origins of Narcotic Control"*, New York: Oxford University Press.

Rydell, C.P. and Everingham, S.S. (1994). „Controlling Cocaine. Supply Versus Demand Programs", *MR-331-ONDCP/A/DPRC*, Santa Monica, CA: RAND.

Rydell, C.P., Caulkins, J.P., and Everingham, S.S. (1996). „Enforcement or Treatment: Modeling the Relative Efficacy of Alternatives for Controlling Cocaine," *Operations Research*, Vol. 44, No. 6, pp.687-695.

Stares, P.B. (1996). „*Global Habit: The Drug Problem in a Borderless World"*. Washington DC: Brookings Institute.

Strogatz, S.H. (1994). „Nonlinear Dynamics and Chaos". Addison-Wesley.

Appendix

A.1 Local Stability Analysis of System *(2)*

Since the initial conditions are always positive (L_0, $H_0>0$) and the first quadrant is an invariant set we restrict our analysis to the positive quadrant of the *(L,H)*-plane. First we derive the equations for the isoclines:

$$\dot{H} = 0: \quad H(L) = \frac{b}{g} L \tag{A1}$$

$$\dot{L} = 0: \quad H(L) = -\frac{L}{q} \ln\left[\frac{a+b}{s} - \frac{t}{sL}\right], \text{ defined for } \frac{t}{a+b} < L \tag{A2}$$

Since $t > 0$, the intersection of the isoclines uniquely defines the equilibrium

$$\hat{E} = \begin{pmatrix} \hat{L} \\ \hat{H} \end{pmatrix} = \frac{t}{\Omega} \begin{pmatrix} 1 \\ b/g \end{pmatrix} \tag{A3}$$

inside a subsection of the positive quadrant of the *(L,H)*-plane ($t/(a+b) < L,\ 0 < H$) if and only if $\Omega := a + b - s \exp[-qb/g] > 0$. Equation (A3) also reflects the equilibrium ratio $L : H = 1 : b/g$, which leads to a equilibrium fraction of heavy users of $b/(b+g)$. The local stability behavior of system *(2)* can be determined in the usual manner by linearization around $\hat{E}$ using the Jacobian matrix evaluated at the stationary point,

$$J|_{\hat{E}} = \begin{pmatrix} \left(sq\frac{b}{g}\right)\exp\left[-q\frac{b}{g}\right] - \Omega & -sq\exp\left[-q\frac{b}{g}\right] \\ b & -g \end{pmatrix}.$$

The corresponding characteristic roots are calculated as follows:

$$l_{1,2} = \frac{1}{2}\left\{\left(sq\frac{b}{g}\right)\exp\left[-q\frac{b}{g}\right] - g - \Omega \pm \sqrt{\left(\left(sq\frac{b}{g}\right)\exp\left[-q\frac{b}{g}\right] - g - \Omega\right)^2 - 4g\Omega}\right\}. \tag{A4}$$

A.2 The Hopf Bifurcation Theorem (Guckenheimer and Holmes, 1983)

Suppose that a family of time-continuos dynamical systems $\dot{z} = F(z, \mu)$, F continuously differentiable, $z \in \Re^n$ and $\mu \in \Re$, has an equilibrium $\hat{z}_c$ at the value μ_c of the parameter, at which the following properties are satisfied:

(i) The Jacobian evaluated at $(\hat{z}_c, \mu_c)$,

$$J(\mu_c) := \left. \frac{\partial F(z, \mu)}{\partial z} \right|_{\mu=\mu_c, z=\hat{z}_c}, \tag{A5}$$

has eigenvalues $\lambda(\mu)$ that vary smoothly with μ and has at $\mu = \mu_c$ a simple pair of purely imaginary eigenvalues, say $\lambda_{1,2} = \pm \omega i$, and no other eigenvalues with real parts zero; μ_c is called a critical value of the bifurcation parameter μ or, for short, the bifurcation point. Then there exists a smooth curve of equilibria depending on the bifurcation parameter μ, denoted $\hat{z}(\mu)$ so that $\hat{z}(\mu_c) = \hat{z}_c$.

(ii) If, moreover,

$$\left. \frac{d(\mathrm{Re}\,\lambda(\mu))}{d\mu} \right|_{\mu=\mu_c} = D \neq 0 \tag{A6}$$

then there is a unique three-dimensional center manifold passing through $(\hat{z}_c, \mu_c)$ in $\Re^n \times \Re$ and a smooth system of coordinates (preserving the planes $\mu = const$) for which the Taylor expansion of degree 3 on the center manifold is given by the following normal form

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} D(\mu - \mu_c) & -(\omega + C(\mu - \mu_c)) \\ \omega + C(\mu - \mu_c) & D(\mu - \mu_c) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \begin{pmatrix} x(x^2 + y^2) \\ y(x^2 + y^2) \end{pmatrix}. \tag{A7}$$

If $A \neq 0$, there is a surface of periodic solutions in the center manifold which has quadratic tangency with the eigenspace of $\lambda_{1,2}(\mu_c)$ agreeing to second order with the paraboloid $\mu = -\frac{A}{D}(x^2 + y^2)$.

(iii) If $A<0$, then these periodic solutions are stable limit cycles, while if $A>0$, the periodic solutions are repelling.

A.3 Evidence of Supercritical Hopf Bifurcation for any of the System-Parameters ν (ν = *a, s, q, b, g*)

At any Hopf bifurcation point $\left(\hat{E}_c, \nu_c\right)$ ($\hat{E}_c$ is determined by Equation (A3)) the normal form of model *(2)* is given by the following system, where $\lambda_{1,2}(\nu_c) = \pm i\omega_c$ denotes the corresponding pair of purely imaginary eigenvalues[6]

$$\begin{pmatrix}\dot{L}\\ \dot{H}\end{pmatrix} = \begin{pmatrix}0 & -\omega_c\\ \omega_c & 0\end{pmatrix}\begin{pmatrix}L-\hat{L}_c\\ H-\hat{H}_c\end{pmatrix} + \begin{pmatrix}\varphi(L,H)\\ \psi(L,H)\end{pmatrix}, \text{ where } \begin{pmatrix}\varphi(L,H)\\ \psi(L,H)\end{pmatrix} = \begin{pmatrix}I(L,H)-(a+b)L+\omega_c\left(H-\hat{H}_c\right)\\ bL-gH-\omega_c\left(L-\hat{L}_c\right)\end{pmatrix}$$

with $\varphi\left(\hat{E}_c\right) = \psi\left(\hat{E}_c\right) = 0$ and $\frac{d}{dt}\varphi\left(\hat{E}_c\right) = \frac{d}{dt}\psi\left(\hat{E}_c\right) = 0$. As long as assumption (A8) holds at $\left(\hat{E}_c, \nu_c\right)$, where accordingly Conditions (A5) and (A6) are satisfied,

$$3g - qb > 0, \tag{A8}$$

the normal form calculation yields a negative cubic coefficient A (for the two-dimensional system[7])

$$A = \frac{1}{16}\underbrace{\left[\underbrace{I_{LLL}}_{<0} + \underbrace{I_{LHH}}_{<0}\right]}_{<0} + \underbrace{\frac{1}{16\omega_g}}_{>0}\underbrace{\left[\underbrace{I_{LH}}_{<0}\left(\underbrace{I_{LL}}_{>0} + \underbrace{I_{HH}}_{>0}\right)\right]}_{<0} < 0.$$

Since Assumption (A8) is fulfilled for the base case set of parameters *a=0.163, b=0.024, g=0.062, s=0.610, q=7.0,* τ *=50,000* and the Hopf critical values (listed in Table 1), respectively, there exists a surface of periodic solutions in the center manifold which has quadratic tangency with the eigenspace of $\lambda_{1,2}(\nu_c)$ for any of the system parameters ν. These periodic solutions are stable limit cycles where the following rules apply (see Strogatz, 1994, p.251):

- The size of the stable limit cycle grows continuously from zero, and increases proportional to $\sqrt{\nu - \nu_c}$, for ν close to ν_c.
- The frequency of the limit cycle is given approximately by ω_c. This formula is exact at the birth of the limit cycle, and correct within $O(\nu - \nu_c)$ for ν close to ν_c. Therefore the period T of the cycle is given by $T = \frac{2\pi}{\omega_c} + O(\nu - \nu_c)$. (A9)

---

[6] $\omega_c = \sqrt{g\left(a+b-s\exp\left[-qb/g\right]\right)}$

[7] $A = \frac{1}{16}\left[\varphi_{LLL} + \varphi_{LHH} + \psi_{LLH} + \psi_{HHH}\right] + \frac{1}{16\omega_c}\left[\varphi_{LH}\left(\varphi_{LL} + \varphi_{HH}\right) - \psi_{LH}\left(\psi_{LL} + \psi_{HH}\right) - \varphi_{LL}\psi_{LL} + \varphi_{HH}\psi_{HH}\right]$

For the definition of the cubic coefficient for higher-dimensional systems see Guckenheimer and Holmes, 1983, p.152ff.

A.4 Calculation of Δ

Rydell and Everingham (1994) estimate that historically 31% of heavy users in the US received treatment and 13% of those treated left heavy use as a result of that treatment. Hence, the outflow rate can be viewed as being

$$g = 0.062 + 0.13\left(u - 0.31\right), \tag{A10}$$

where u is the proportion of heavy users receiving treatment each year. Assume that we increase that proportion $u \to \tilde{u}$ and the outflow rate $g \to \tilde{g}$, respectively, according to Equation (A10). We define total discounted consumption, for a fixed outflow rate g, over a relevant finite planning horizon such as 50 years, starting with any combination of light and heavy users,

$$Q_{50_g} := \int_0^{50} e^{-0.04\,t}\left(16.42\,L_g(t) + 118.93\,H_g(t)\right)dt$$

(Following Everingham and Rydell, we assume light and heavy users consume at an annual rate of 16.42 grams and 118.93 grams per year, respectively, and discount at 4% per annum) and we consider the change in discounted consumption in grams due to additional treatment, denoted by $Q_{50_g} - Q_{50_{\tilde{g}}}$.

The discounted cost of additional treatment ($u \to \tilde{u}$) over the same finite planing horizon is calculated as follows

$$Cost_{50_g} = C_{treat} \underbrace{\int_0^{50} e^{-0.04\,t}\,uH_g(t)\,dt}_{:= C_{50_g}}$$

where C_{treat} reflects the cost of a heavy user's treatment. To achieve the increment in the proportion of heavy users receiving treatment each year, $u \to \tilde{u}$, and the outflow rate $g \to \tilde{g}$, respectively, one has to provide an additional treatment budget, $C_{treat}\left(C_{50_{\tilde{g}}} - C_{50_g}\right)$. Therefore Δ may be interpreted as the averted consumption in grams per additionally treated heavy user

$$\mathsf{d} = \frac{\mathsf{k}}{C_{treat}} \underbrace{\left(\frac{Q_{50_g} - Q_{50_{\tilde{g}}}}{C_{50_{\tilde{g}}} - C_{50_g}}\right)}_{:= \Delta} \tag{A11}$$

where k measures the reduction in social cost per gram of consumption averted. Then a policy maker would have to estimate the cost of treating a user, C_{treat}, and judge the benefit of reducing cocaine consumption by a certain amount Δ. Condition (A11) determines how many grams of consumption an additional treatment must avert in order for additional treatment to be cost-effective ($\mathsf{d} \geq 1$).