

# Understanding Inertia: Inherent Limitations on Evaluating “Upstream” Prevention Interventions

**Short Title:** Understanding Inertia

Jonathan P. Caulkins

Carnegie Mellon University  
Qatar Campus and H. John Heinz III School of Public Policy & Management  
5000 Forbes Ave. Pittsburgh, PA 15213

[caulkins@cmu.edu](mailto:caulkins@cmu.edu)

1-412-268-9590 (office)

1-412-268-5338 (FAX)

Irmgard Zeiler

Department of Mathematical Methods in Economics, Vienna University of  
Technology, Argentinierstrasse 8/119, A-1040 Vienna, Austria

[Irmgard.Zeiler@gmx.net](mailto:Irmgard.Zeiler@gmx.net)

Carnegie Mellon University  
Qatar Campus and H. John Heinz III School of Public Policy & Management  
5000 Forbes Ave. Pittsburgh, PA 15213

March, 2008

## **Acknowledgments**

This work was funded in part by the Robert Wood Johnson Foundation, the Qatar Foundation, and the Colonial Foundation through its support of Australia’s Drug Policy Modeling Program.

# Understanding Inertia: Inherent Limitations on Evaluating “Upstream” Prevention Interventions

## **Abstract**

When different types of policy interventions are available, there is an understandable desire to evaluate all alternatives using common metrics so scarce resources can be allocated in the most efficient manner. However, systems that display significant lags in their response to some interventions can confound such an empirical approach. This paper provides a parsimonious mathematical representation of some of the challenges confronted when trying to evaluate upstream interventions on lagged systems to help clarify when it is and when it is not practical to expect those interventions to meet the same standard of proof as downstream interventions. Implications for drug policy and delinquency prevention are elaborated.

*(Keywords: Prevention, drug policy, evaluation, cost-effectiveness, policy modeling)*

## **1. Introduction**

Many types of anti-social behavior are usefully thought of as progressing through a “career” or “life-course trajectory” (Farrington, 2003; Nagin, 2005; Nagin and Tremblay, 2005). Initiation typically occurs during adolescence since adults who have not previously engaged in the activity are unlikely to start. Most people “quit” or “mature out” after only modest participation, but others escalate to more intense forms of behavior that can persist for years and generate substantial societal costs. This general pattern holds for criminal offending (Blumstein et al., 1986; Sampson and Laub, 2005), drug use (Makkai and Payne, 2003; Hser et al., 1997, 2007a, 2007b), drug selling (Johnson et al., 1995), and smoking (Graham et al., 2006). For the sake of concreteness, we present data and analysis for control of illegal drug use, but many of the qualitative insights carry over to other domains.

Two general strategies for controlling the resulting societal costs can be distinguished. “Upstream” interventions reduce initiation and/or escalation from lower-level participation to more serious offending while “downstream” interventions focus on current “high-rate” offenders. Upstream interventions are typically cheaper per person, but they may not have high efficacy, can take time to make a difference, and often have difficulty distinguishing between people who are merely dabbling in delinquency and those who would otherwise go on to be high-rate offenders. Downstream interventions target directly the behaviors that are most problematic, but may be expensive and/or not very effective.

Downstream drug control interventions include treatment, incarcerating users, and enforcement/treatment partnerships such as drug courts or coerced abstinence. Upstream interventions include school-based prevention, community-based prevention, and disrupting supply in ways that suppress initiation for the duration of the disruption.

We have been extensively involved over many years in analysis and policy debate concerning the relative effectiveness of various drug, crime, and tobacco control interventions. This paper is motivated by our observation that sometimes demands for rigorous, evidence-based policy can turn subtly biased against upstream interventions because the consequences of system lags are not fully appreciated. The three sections of this paper offer (1) a numerical example that illustrates the issue, (2) a concise mathematical representation of system inertia, and (3) a structural explanation for why so many systems involving anti-social behavior display this sort of inertia. A discussion of implications follows.

Two points of clarification are in order. First, we are not disputing that some upstream interventions have no discernable effect on any outcome; such interventions are of no value regardless of whether the system displays inertia or not. Rather, we are suggesting that even the subset of interventions that do successfully affect upstream behavior might still have a hard time impressing observers focused on “bottom-line” outcomes if those outcomes are primarily measured downstream.

Second, this argument pertains to any type of upstream intervention; it is not a defense of prevention programs specifically. We intentionally mentioned above both demand-side and law enforcement interventions in our lists of both upstream and

downstream drug control interventions. Indeed, the genesis of this paper was thinking about how to evaluate drug market supply disruptions.

So throughout the body of the paper we will discuss the ability to observe the effects of a hypothetical, unspecified intervention that has a dramatic effect on initiation. In the discussion we return to the issue of differential implications for prevention programming vs. supply disruption or other interventions.

## **2. Empirical Illustration**

We begin with a numerical simulation using data from the US National Household Survey on Drug Use and Health.<sup>1</sup> Figure 1 shows as a function of time since initiation the proportion of respondents to the 1999 – 2005 Surveys who self-reported receiving substance abuse treatment in the last twelve months (variable ILLTRMT) and the proportion judged to be drug dependent based on the standard proxy to the DSM-IV criteria (variable DEPNDILL)<sup>2</sup>. The Figure contains two lines for each outcome: one for time since marijuana initiation; the other for time since initiation of any illegal drug. Results are very similar so in the sequel we will report only the latter.

Figure 1 is interesting in its own right. It suggests that for these respondents, dependence peaked five years after initiation, but then decreased markedly through year 17. Treatment participation had a much less pronounced peak and decline. So the proportion of dependent users in the household population who report recently receiving treatment rose from one-third in the first ten years after initiation to about two-thirds after year 17. Since assessments of the “treatment gap” in the US rely heavily on these same data (OAS, 2002), the figure suggests that most of that assessed gap comprises people who have been using for less than fifteen years.<sup>3</sup>

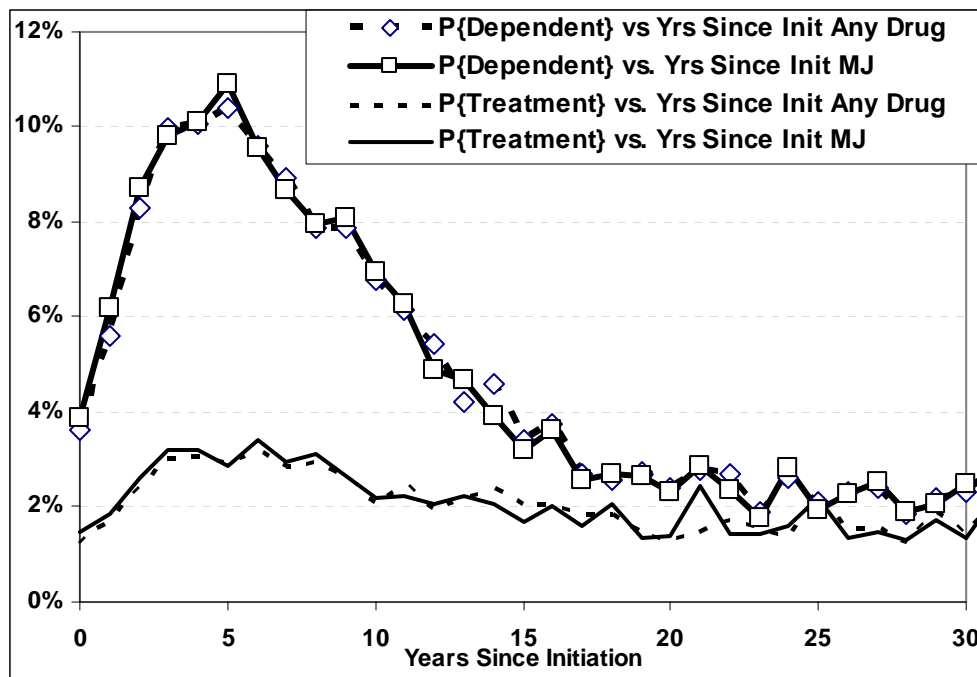
---

<sup>1</sup> See SAMHSA (2006) for a description of this data system. All analyses here were produced via SAMHDA’s on-line data analysis tool available at <http://webapp.icpsr.umich.edu/cgi-bin/SDA/SAMHDA/>.

<sup>2</sup> Presumption of dependence is developed from respondents’ answers to questions such as whether the drug caused them to spend less time on work or other important activities, whether they had tried to cut down but failed, whether they needed more to get the same effect, etc. See SAMHSA (2002) for details.

<sup>3</sup> Note: Figure 1 confounds time and cohort effects. Since the data come from the 1999 – 2005 surveys, the plotting point for people thirty years after initiation comes from people who initiated between 1969 – 1975, whereas the plotting point for outcomes ten years after initiation comes from respondents who reported initiating between 1989 – 1995. That is a potential limitation for treatment gap projections, but for our primary purpose below all we need is some characterization of the likelihood of an outcome as a function of time since initiation. Our conclusions depend only on the

Figure 1: Proportions of Household Survey Respondents Judged to Be Dependent and Who Report Receiving Substance Abuse Treatment in Past Year as a Function of Years Since Initiation into Marijuana and Into Any Drug Use



For present purposes, however, we merely want to use these series to understand what would happen to aggregate measures of dependence and treatment demand among the household population if some upstream intervention completely erased an entire cohort of initiation. The answer depends on how much initiation there was in the past relative to what there would have been in the year for which initiation was eliminated. To keep things simple, suppose that initiation had been perfectly constant since the infinite past, so the system was in steady state, and that after the one year interruption, initiation returns to baseline levels. In other words, initiation had been holding steady, suddenly ground to an absolute halt, then, a year later, resumed its previous level. That is about as abrupt an upstream shock as one can imagine, so one would expect to see a dramatic response in outcome measures.

We can check this by plotting the effect of a one-year interruption as follows. Take a particular point in time, say five years after the interruption. In that year, there are people who are dependent and/or in treatment who initiated 0, 1, 2, 3, and 4 years

---

general shape of the curve, and the general shape is almost certainly of the nature depicted in Figure 1. We need specific plotting points to proceed with our numerical simulation, but it is only the general shape that matters, not specific values.

earlier. There are also people who initiated 6, 7, ..., 30 years earlier.<sup>4</sup> But there are no people who initiated 5 years earlier because, by assumption, that initiation cohort was eliminated by the upstream intervention. We know how many people initiated in each year (a fixed constant in all years except the year when there were no initiates), and Figure 1 tells us what proportion of them would be dependent or in treatment in the year in question. So it is a simple arithmetic exercise to plot how many people are in treatment or are dependent as a proportion of the number who would have been in treatment or dependent had there not been the interruption in initiation. Figure 2 plots the results.

Figure 2: Projected Effects of a One-Year Interruption in Initiation on Drug Dependence and Treatment Among the Household Population; Outcome Expressed As a Proportion of Baseline (No Interruption) Scenario

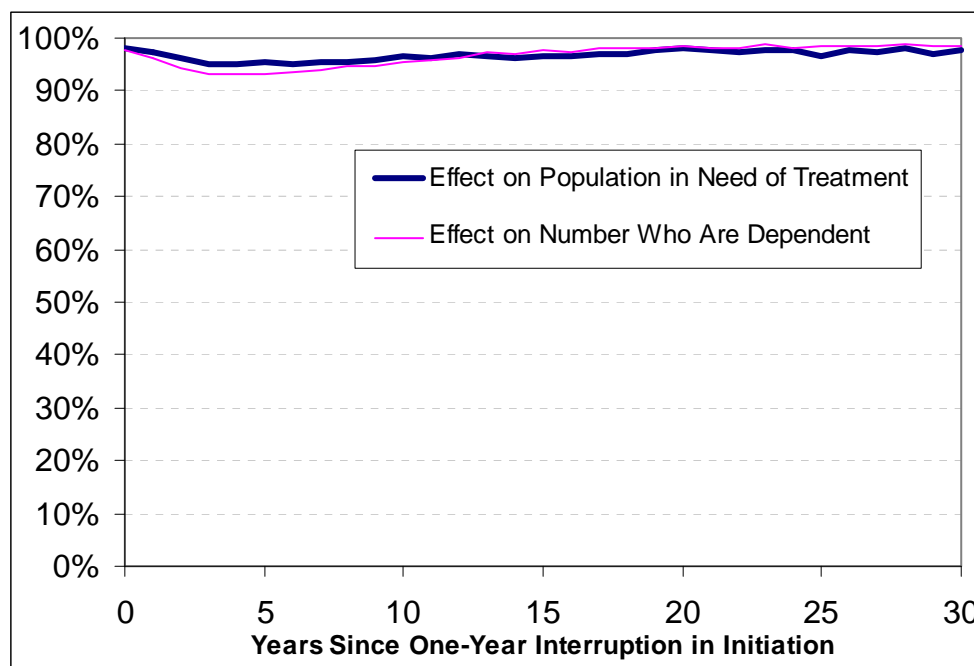


Figure 2 shows that even a complete and total one-year interruption in initiation creates only a modest ripple in measures of drug treatment and dependence. Rates decline by 2% in the first year, falling further to a trough of 5% below baseline around year 6, and rebound to be within 2% of baseline by roughly year 20.

<sup>4</sup> We truncate after 30 years because the sample sizes underpinning the plotting points in Figure 1 get small after year 30. (Initiation into drug use was much lower earlier in the 1960s than in later years.) Also, assuming most people initiate by age 25, truncating after 30 years minimizes selection effects due to premature mortality because mortality is not extremely high for drug users in the household population before age 55

This subdued response is not because a one-year interruption in initiation is of little value. Since the system was assumed to be in steady state and the calculations do not include any feedback or nonlinear effects, eliminating one year of initiation will reduce total drug use over time by exactly one year of steady state consumption, which cost of illness studies value at around \$180 billion in the US (ONDPCP, 2004).

Rather, the effects are hidden in plain sight by being spread out so smoothly over time. It is not that the effects are small, but rather that they are diffused over time.

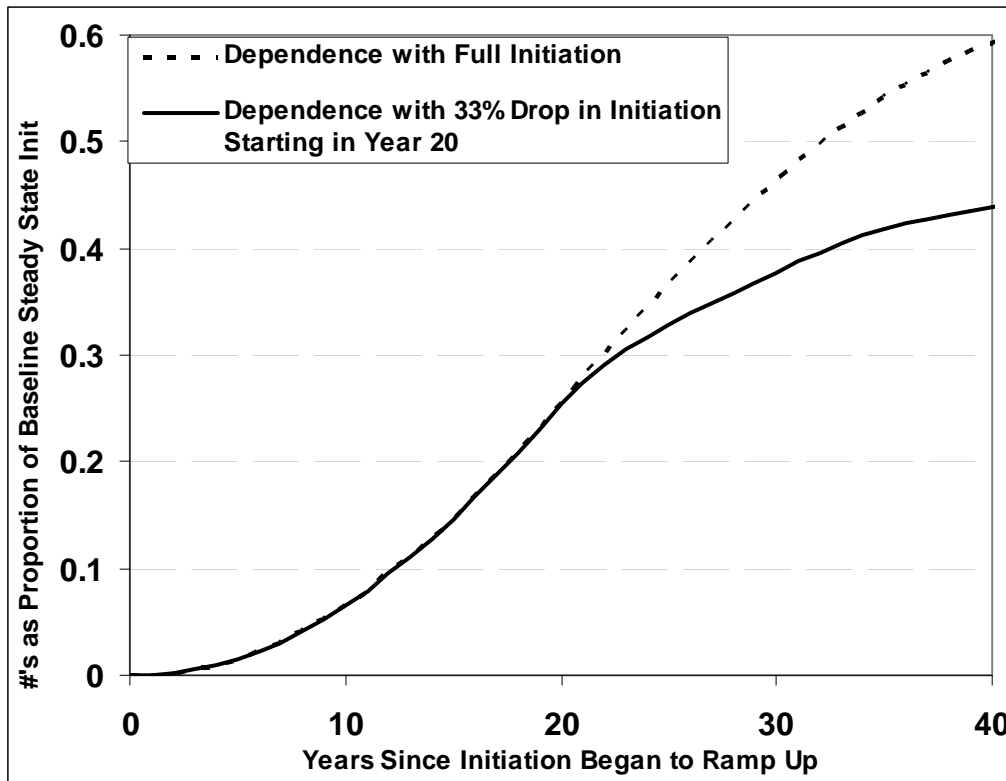
If there were no measurement error, a 2% drop could be detected, but in reality there is considerable measurement error. For Wright's (2004, Table A.15) state-level estimates of drug dependence, the prediction interval was on average half the size of the point-estimate, implying confidence intervals of +/-25% from sampling variability and model bias. If the +/-25% range represents a normally distributed 95% confidence interval, then the odds of a pre-post design recording any decline at all given a true 2% decline are only 1.19:1 and the chance of perceiving the decline to be statistically significant at the  $\alpha = 0.05$  level in a one-tailed test is only 6.3%, barely greater than the value of  $\alpha$  itself.

A second complication is that systems with great inertia take a long time to reach steady state. During the run-up to steady state, it is even harder to see the fingerprint of upstream interventions because the counterfactual would be an upward trend, not stability. Small and temporally diffuse reductions from an upward trend still leave an upward trend, just one with a slightly smaller slope. Unless the counterfactual is known with great precision (which is rarely the case), that means even successful upstream interventions will be followed by adverse trends in aggregate indicators.

Figure 3 illustrates this using the dependence vs. time since initiation curve from Figure 1 and the following initiation scenario. Initiation grows linearly to a peak in 20 years and then either (1) stabilizes (dotted curve) or (2) is cut permanently by one-third (solid curve). Clearly if one knew the counterfactual (the dotted line) it would be easy to see that the one-third cut in initiation was a boon. However, in the more typical circumstance where the counter-factual is not known, all one sees after the dramatically successful upstream intervention is steadily rising need for treatment (solid line that continues to grow).

The next section presents a parsimonious mathematical representation of inertia that both captures and distinguishes these two separate considerations that complicate empirical evaluation of upstream interventions.

Figure 3: In a “New” System Inertia Make Even Permanent Reductions in Initiation Hard to Appreciate; Projected Prevalence of Drug Dependence with and without a Permanent 33% Cut in Initiation Starting in Year 20



### 3. Mathematical Representation

Let

- $f(t)$  = probability an individual manifests the outcome of interest (e.g., drug dependence)  $t$  time units after initiation,
- $I(y)$  = initiation in year  $y$ , and
- $X(t)$  = the aggregate (societal) level of the outcome of interest at time  $t$ .

Then  $X(t)$  is computed via the convolution integral

$$X(t) = \int_{-\infty}^t I(y) f(t-y) dy. \quad (1)$$

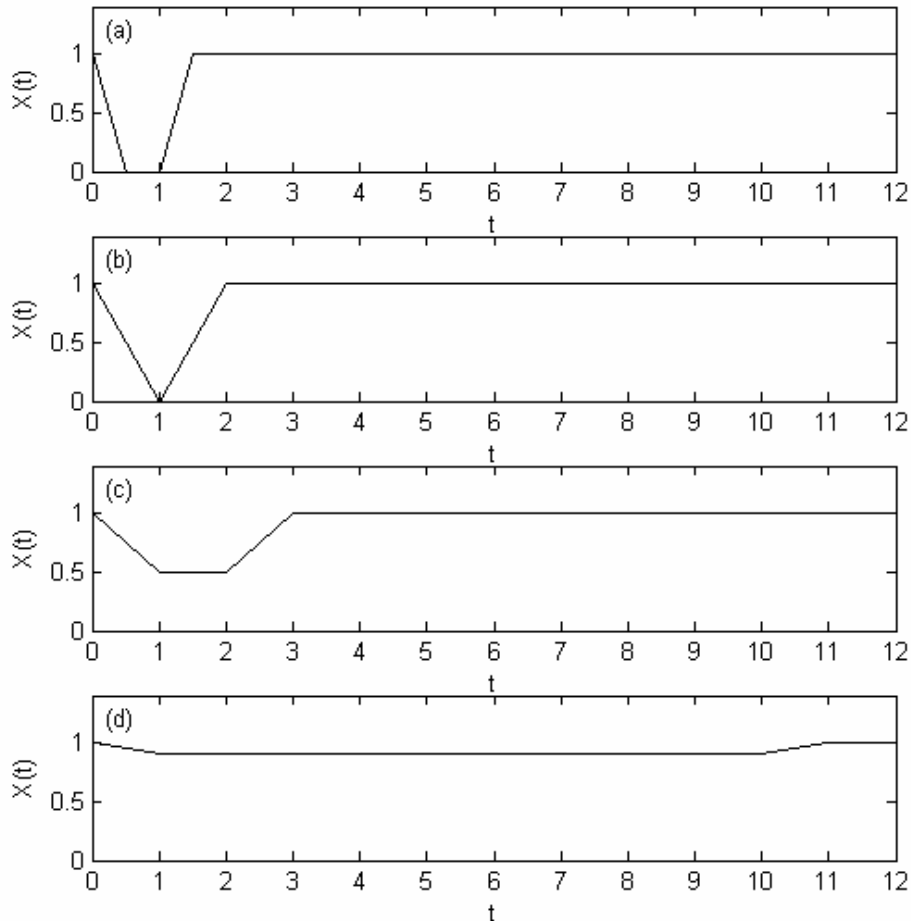
This formula concisely distinguishes the issue of rising initiation  $I(y)$  before the upstream intervention takes effect from the “inertia” embodied in the function  $f(t)$ .

Two attributes of  $f(t)$  can obscure the effect of a temporary reduction in  $I(y)$  on the population-level outcome  $X(t)$ : (1) building slowly over time and (2) being relatively flat, as opposed to having a single short and well-defined spike. The first (lag) creates a delay between the upstream intervention and the resulting effects on  $X(t)$ . The second (smoothing) translates even an abrupt change in initiation into a diffuse effect on  $X(t)$ .

Figure 1 contained plots of  $f(t)$  showing that among household survey respondents the outcomes of dependence and being treated in the past-year manifest both lag and smoothing. There is also considerable smoothing for being a past-year use of some substances (e.g., marijuana) but not others (e.g., LSD). Both delay and smoothing are prominent in a model of Australian drug use that traces trajectories from first initiation through dependent injection drug use (Caulkins, 2008). If one starts not with first drug use but rather with entrance into heroin treatment in the US, there may be no delay but still considerable smoothing (cf., Hser et al., 2007b, Figure 2).

Equation 1 is a very concise representation of a subtle concept, so concise that it may be opaque to readers not familiar with convolution integrals. So before proceeding, consider the following simple example. Suppose that everyone who tries a drug uses steadily for  $T$  years (i.e.,  $f(t) = 1$  for  $0 < t \leq T$ ), that initiation has long remained at some baseline level normalized without loss of generality to  $1/T$ , and that initiation is eliminated for one year starting at time  $t = 0$ . Figure 4 shows the effect on  $X(t)$  for four cases: (a)  $T = 0.5$ , (b)  $T = 1$ , (c)  $T = 2$ , and (d)  $T = 10$ .

Figure 4: Illustration of Convolution Integral: Effect on  $X(t)$  of a one year Disruption to Initiation  $I(y)$  for a Simple Individual-Level Behavior  $f(t) = 1$  for  $0 < t \leq T$ , for (a)  $T = 0.5$ , (b)  $T = 1$ , (c)  $T = 2$ , and (d)  $T = 10$



In case (a), since people use the drug for only half a year, six months after initiation stops no one is using the drug. So at  $t = 0.5$ ,  $X(t)$  falls to zero. It remains at 0 until the end of the one-year interruption in initiation. By six months after that ( $t = 1.5$ ), prevalence is back to its baseline level because with career lengths of only 0.5 years, nothing that occurs before  $t = 1$  affects the prevalence at  $t = 1.5$ . The resulting  $X(t)$  graph clearly shows the effect of the interruption in initiation.

Case (b) is similar, but now people use the drug for one full year ( $T = 1$ ), so it takes the full one-year interruption in initiation before prevalence falls to 0, and prevalence immediately picks up again with the resumed flow of initiates. That is, when  $T = 1$ , a one-year interruption creates just one fleeting moment when prevalence is zero. Such an outcome would still be very noticeable.

In case (c) prevalence never reaches zero. Even at the end of the one-year interruption to initiation, prevalence is still half its baseline level because all of the people who initiated between 24 and 12 months ago are still using. For a year there is no change in prevalence even though initiation has rebounded to its initial level. E.g., current users at  $t = 1.5$  include half a year's worth of initiates who started between 24 and 18 months before plus another half a year's worth of initiates who started 0 to 6 months earlier. Finally, two years after the beginning of the disruption, prevalence begins to rise because the effects of the early part of the disruption can no longer be felt. By two years after the end of the disruption, i.e., at  $t = 3$ , no part of the disruption any longer affects current prevalence, so  $X(t)$  has rebounded to baseline levels.

The same logic applies in case (d), where  $T = 10$ . Prevalence declines linearly between  $t = 0$  and  $t = 1$ . At  $t = 1$ , prevalence is 10% below baseline levels. That level is maintained until  $T = 10$  years after the start of the disruption (i.e., until  $t = 10$ ) and baseline prevalence is restored  $T = 10$  years after the end of the disruption.

The upshot of Figure 4 is clear. When  $f(t)$  is positive for a long time relative to the duration of a disruption to initiation, then the effects of the disruption are hard to see in a plot of the downstream measure,  $X(t)$ , which was prevalence in the example of Figure 4.

Lags in  $f(t)$  shift to the right the “bite” that disrupting initiation takes out of  $X(t)$ . When the lag exceeds the length of the disruption, there will be no correlation between program effects and the outcome unless a properly lagged evaluation model is used.

In conclusion, the flatter and more lagged the function  $f(t)$ , the more obscure or distorted is the “signal” of the initiation disruption manifested in  $X(t)$ . Conversely, the briefer and sharper the  $f(t)$ , the easier it is to see the signal. Indeed, in the limiting case where  $f(t)$  reduces to the Dirac delta function, then  $X(t)$  is identical to  $I(t)$  and one can observe the intervention's effect with no distortion.

### **3. Inertia in Sequential, Multi-Stage Models of Career Progression**

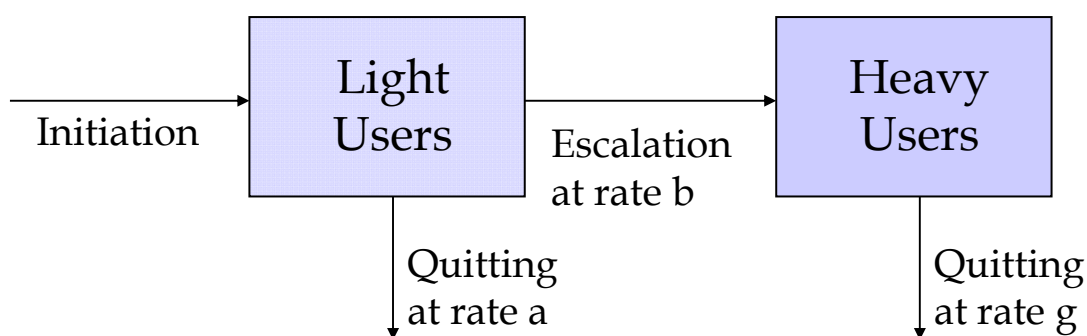
#### *3.1 Two-Stage “LH” Model of Delinquency*

The previous section introduced a concise mathematical expression that explains why system inertia makes it hard to assess upstream interventions by looking

at downstream indicators. In short, lagged and smoothed  $f(t)$  functions camouflage the effects of fluctuations in initiation,  $I(y)$ . That begs the question of why so many types of delinquency are characterized by  $f(t)$  functions that are lagged and smooth? The answer is that this is a direct consequence of the basic problem descriptors set out in the introduction. If people start with a low-intensity of behavior, most soon “mature out”, and only a subset escalate to more intense forms of behavior that can persist for years and generate substantial societal costs, then the resulting  $f(t)$  will be lagged and relatively flat.

These broad statements have long been known to apply to drug use trajectories (Chen and Kandel, 1995). Everingham & Rydell (1994) translated them into a simple Markov model that might be described as the classic elementary model of drug-use careers. Everingham & Rydell differentiated just two intensities of drug use ( $L$  for “light” and  $H$  for “heavy”), with all users starting in the  $L$  state. Most soon quit (at a constant per capita rate  $a$ ) without progressing, but a smaller number escalate (at rate  $b$ ) to the  $H$  state, from which they quit at a much lower rate,  $g$ . Note: Quitting from the  $H$  state should be understood as being net of relapse. Interruptions of use are common, whether driven externally (e.g., by incarceration) or internally (ceasing use for a time). So the  $H$  state includes people who are temporarily not using, but who will use again at high rates. This “ $LH$ ” model is depicted graphically in Figure 5.

Figure 5:  $LH$  Model of Drug Use Trajectories



Fancier models with more states have been created, but the simple  $LH$  model with just three parameters describes well the thirty year history of the modern US cocaine epidemic (Caulkins et al., 2004), and has been the basis for a variety of extensions (e.g., Behrens et al., 1999, 2000). Furthermore, when interpreted as a

Markov process, its feed-forward structure (heavy users never return to being indistinguishable from people who have never escalated) allows one to write an explicit expression for the probability of being in a state given that one started at time  $t = 0$  in the  $L$  state:

$$L(t) = e^{-(a+b)t} \quad \text{and} \quad (2)$$

$$H(t) = \frac{b}{a+b-g} \left( e^{-gt} - e^{-(a+b)t} \right).$$

Most drug-related harms are generated by heavy or dependent users. Indeed, heavy users consume so much more per capita than do light users that beyond the early explosive growth stage of a drug epidemic, even most drug use (and, hence, drug market demand) comes from heavy users (Caulkins et al., 2004). So a not unreasonable simplification would think of “downstream indicators” of drug-related problems as being proportional to the number of people in the  $H$  state. Thus, in terms of Equation 1,

$$f(t) = H(t) = \frac{b}{a+b-g} \left( e^{-gt} - e^{-(a+b)t} \right), \quad (3)$$

although one might also consider a weighted sum of light and heavy users, weighting by their relative rates of participation (cf., Behrens et al., 2000).

The basic shape of function  $H(t)$  holds for any set of parameters values with  $a+b > g$  (which follows since exit from heavy use,  $g$ , is slower than exit from light use,  $a$ ). Initially, at  $t = 0$ , both  $e^{-gt}$  and  $e^{-(a+b)t}$  equal 1, so  $f(0) = 0$ . The function  $f(t)$  increases to a maximum at  $t^{peak} = \ln((a+b)/g)/(a+b-g)$ . The average times spent in the  $L$  and  $H$  states (dwell times) conditional on entering are  $1/(a+b)$  and  $1/g$ , respectively, so

$$t^{peak} = \frac{\ln((a+b)/g)}{a+b-g} > \frac{\ln((a+b)/g)}{a+b} = \text{dwell time in } L * \ln\left(\frac{\text{dwell time in } H}{\text{dwell time in } L}\right). \quad (4)$$

For example, if individuals typically spend three years in the low-intensity state and fifteen in the high-intensity state once they get there, then the peak in  $f(t)$  does not occur until more than five years after initiation.

Viewing a scaled version of  $f(t)$  as a distribution of delays between initiation and a randomly selected year of heavy use, the average and standard deviation of the delay between initiation and heavy use are

$$t^{avg} = \frac{1}{a+b} + \frac{1}{g}$$

$$\sigma_t = \sqrt{\frac{1}{(a+b)^2} + \frac{1}{g^2}} \quad (5)$$

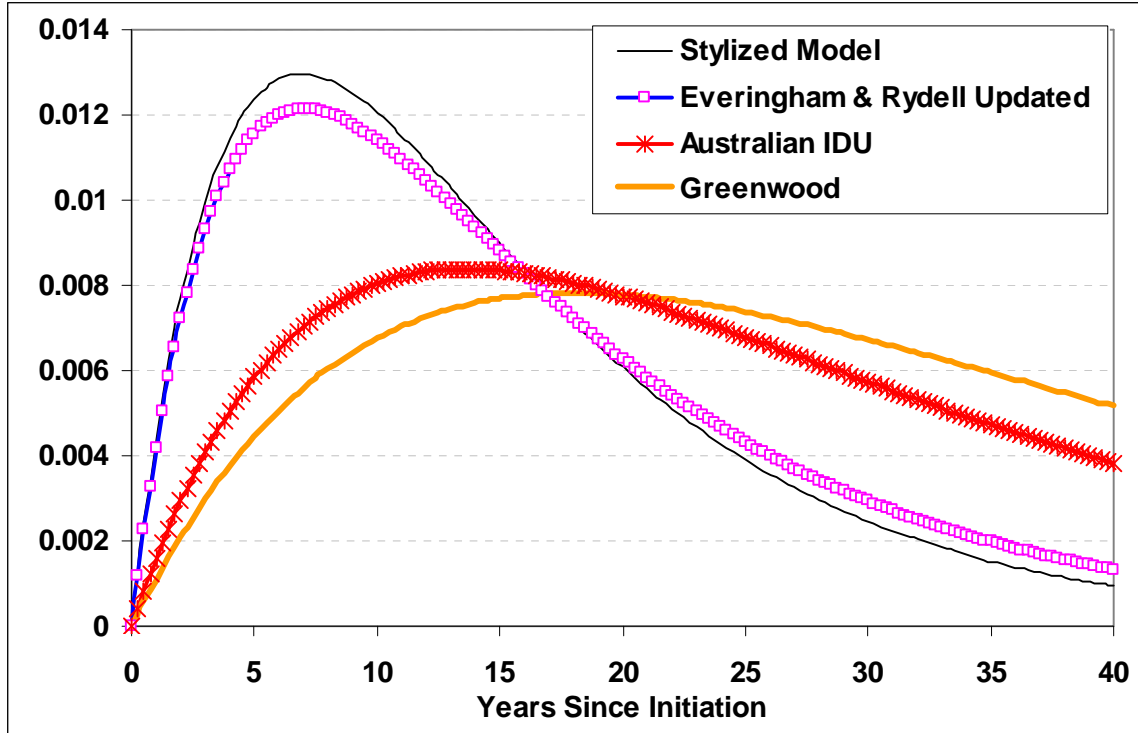
Their ratio, the coefficient of variation (CV), is a measure of how spread out the function  $f(t)$  is. Table 1 shows the  $a$ ,  $b$ , and  $g$  values and implied  $t^{peak}$ ,  $t^{avg}$ , and CV for four models: (1) the Everingham & Rydell model of US cocaine demand with parameters as updated by Caulkins et al. (2004), (2) the Greenwood et al. (1994) model of low- and high-rate criminal offenders, (3) the Caulkins et al. (2007) model of Australian injection drug use simplified to two states, and (4) a generic two-state delinquency model which follows the Pareto Law in the sense that the 20% of offenders who escalate to the  $H$ -state are responsible for 80% of offending because they persist in that state so long (10 years) and commit offenses at a much higher rate (10 times higher).

Table 1: Four Parameterizations of the  $LH$  Model Structure

Parameter	Everingham & Rydell, updated by Caulkins et al. (2004)	Greenwood et al. (1994)	Australia n IDU	Stylized Model
$a$	0.2	0.049	0.096	0.16
$b$	0.03	0.012	0.007	0.04
$g$	0.08	0.05	0.05	0.1
$t^{peak}$	7.0	18.1	13.6	6.9
$t^{avg}$	16.8	36.4	29.7	15.0
CV of lag between initiation and heavy use	1.28	2.22	1.06	1.45

Figure 6 plots the  $f(t)$  curves for these models. The curves for the stylized model and updated Everingham & Rydell model have inertia comparable to the dependence curve in Figure 1. The Australian IDU and Greenwood parameterizations display even greater inertia.

Figure 6:  $f(t)$  for the Four  $LH$  Model Parameterizations in Table 1, Scaled So Each Integrates to 1.0 Over 40 Years



### 3.2 Multi-Stage Models

The analysis can be extended to models with more than two stages of progression. For a three-state system, the formulas are only moderately more complex. In particular, if we add a third box to differentiate  $M$  = medium frequency users from  $L$  = light and  $H$  = heavy users and let  $c$  and  $d$  denote the quit and escalation rates from this new state so that

$$\begin{aligned}
 \dot{L} &= -(a+b)L, \\
 \dot{M} &= bL - cM - dM, \\
 \dot{H} &= dM - gH, \\
 L(0) &= 1, M(0) = 0, H(0) = 0.
 \end{aligned} \tag{6}$$

then  $H(t)$  is given by

$$H(t) = \frac{db(e^{-(c+g)t}(a+b-g) + e^{-(a+b)t}(e^{-(c+g)t}(-a-b+c+g) - c-d))}{(a+b-c-d)(a+b-g)(c+d-g)}. \tag{7}$$

Except in certain special cases the expressions for  $H(t)$  become harder to interpret as more states are added. One such case is when every state has the same

total exit rate so that the quit rate plus the escalation rates sum to a constant or, equivalently, the states all have the same average dwell time. In that case, if the constant  $m$  denotes this common exit rate and  $X_i$  denotes stage  $i$  (e.g.,  $i = 0$  for  $L$ , 1 for  $M$ , and 2 for  $H$ ), then the three-state system can be rewritten as:

$$\begin{aligned}
\dot{X}_0 &= -m X_0 \\
\dot{X}_1 &= m_1 X_0 - m X_1 \\
\dot{X}_2 &= m_2 X_1 - m X_2 \\
X_0(0) &= 1, X_1(0) = 0, X_2(0) = 0.
\end{aligned} \tag{8}$$

The solution of which is given by

$$\begin{aligned}
X_0(t) &= e^{-mt} \\
X_1(t) &= m_1 t e^{-mt} \\
X_2(t) &= \frac{m_1 m_2}{2} t^2 e^{-mt}
\end{aligned} \tag{9}$$

Generalizing this to an  $(n+1)$ -stage system with initiation into stage 0, the explicit solution of the final stage  $X_n(t)$  is given by

$$X_n(t) = \frac{t^n}{n!} e^{-mt} \prod_{i=1}^n m_i. \tag{10}$$

This function has a peak at  $t^{peak} = n/m$ , a mean at  $t^{avg} = (n+1)/m$ , and – when normalized to have unit area – a standard deviation of  $\sigma = \sqrt{n+1}/m$ . So the more stages, the later and more spread out is the resulting  $f(t)$ , i.e., the greater the inertia.

Thus the two-state case elaborated above is in some sense a best case in terms of ability to detect upstream intervention's effects on downstream indicators. Yet, as the numerical examples showed, even in that best case, observers focused on downstream indicators are unlikely to appreciate even successful upstream interventions.

#### 4. Discussion

The discussion above is a warning not to draw inferences about the effectiveness of upstream drug control interventions based on effects – or lack thereof – on downstream indicators. Doing so would be akin to trying to assess the effectiveness of HIV prevention strategies by looking at changes in AIDS prevalence, which is difficult because of AIDS' long incubation period.

That there is an analogy with another domain, such as HIV/AIDS, is not surprising. Although we made the argument using examples and data from drug control, we claimed in the introduction that the arguments also apply to efforts to control crime, violence, and use of legal addictive substances. Indeed, they ought to apply to any type of risky behavior for which only a subset of people escalate to the dependent/high-frequency/problematic end of the spectrum of behaviors and those that do reach that more problematic state remain there for an extended period. The key is not that the behavior involves drugs or delinquency more generally, but rather that the “systems” involve considerable lags and inertia.

Many observers focus on downstream indicators. That is understandable since most social costs come from downstream behaviors. The central message of this paper is, those observers should not presume upstream interventions are ineffective just because they do not visibly affect downstream indicators. No news is no news in this case.

Avoiding mistaken inference is valuable, but it begs the question, “How then should upstream interventions be evaluated?” There are at least three broad possibilities.

The first is to evaluate upstream interventions only with respect to upstream indicators and not to compare them to downstream interventions that affect downstream indicators. That is, one could simply give up on the idea of predicating strategic resource allocation between upstream and downstream interventions on quantitative evaluations conducted using common metrics. To give that up is giving up a lot; it means giving up a foundational premise for efficiency and good governance. Nevertheless, as a practical matter, policy has been made in quite a few domains for quite some time without benefit of explicitly comparing upstream and downstream interventions on common metrics.

The second option is to translate upstream and downstream outcomes into common metrics. For example, a benefit-cost analysis might try to monetize the proximate non-monetary outcomes of both upstream and downstream interventions. Superficially this seems like a great idea. In reality, if not done and interpreted properly, it can be strongly biased against upstream interventions because it appears to judge them on a fair playing field when in fact it does not. The reason is simply that most of the economic value of upstream interventions can come from their effects

on downstream indicators, even if those effects are hard to see against background noise because they are delayed and diffuse. Upstream interventions' effects on upstream indicators (e.g., changes in the amount of light drug use or low-level offending) may not generate much economic value because those upstream behaviors are associated with few social costs.

The third alternative is to apply the benefit-cost analysis or monetization not (only) to proximate outcomes, but instead to first project or model effects on downstream indicators by combining observed effects on upstream indicators with some assumed causal model that relates changes in upstream indicators to projected, subsequent downstream effects. Advantages of this approach are clear. It holds the potential for comparing upstream and downstream interventions on a common metric that is not inherently biased against upstream interventions. Also, the analysis can be done as soon as proximate effects are observed; there is no need to wait 20 years to consider effects that are not expected to occur until 20 years after the intervention. Furthermore, if the model projects outcomes for the same indicators that are used to evaluate directly the downstream interventions, one can sidestep altogether the challenges of monetizing outcomes by making the comparison a cost-effectiveness rather than benefit-cost analysis. (For those with great confidence in economists' abilities to scientifically monetize diverse outcomes that is no advantage at all, but some audiences are more skeptical of that monetization.)

Against those advantages are some obvious disadvantages. Notably, model-based projections require a model. Such models may not exist at all and, even if they do, will be predicated on a host of assumptions, many if not most of which cannot be validated.

How do these advantages and disadvantages stack up for specific types of upstream interventions? The most-studied upstream interventions are individual and classroom-based interventions. They lend themselves to rigorous evaluations because these are planned interventions whose unit of analysis makes randomized control trials feasible. Hence, the first type of evaluation (evaluating solely in terms of observed upstream impacts) is possible and, in fact, is the norm. Model-based analyses are also possible because life-course trajectory or "career" modeling paradigms provide a basis for making projections of long-run effects on downstream indicators.

Evaluation opportunities are much more limited for community-level interventions, whether they are media campaigns, red-ribbon weeks, or general investments in economic development. In those cases, it is hard to identify control groups, so even proximate effects are hard to establish scientifically, and without sound estimation of proximate effects, no monetization or model-based projections of downstream effects are reliable, no matter how good the monetization strategy or model.

A third type of upstream intervention is an abrupt shock to the system at the community level. The prototypical example would be a disruption to drug supply that lasts at least a year or two (e.g., Crane et al., 1997; Weatherburn et al., 2003). Such outcomes can't be planned, so prospective studies are not possible, and control groups are no more available than for planned community-level interventions. However, sometimes the shocks are so dramatic (e.g., an 80% drop in heroin overdose in Victoria after the Australian heroin drought, Moore et al. 2005) that an interrupted time series analysis might give reasonably sound estimates of proximate effects. Furthermore, economics gives a framework for projecting downstream consequences (e.g., the notion of elasticity of demand can translate effects on prices into effects on initiation and consumption).

In short, there may not be one evaluation recipe that is the best for all types of upstream interventions. To the extent that statement is true, it suggests another important caution. Policy makers impatient for an empirical or evidence-base for their decision making can be tempted to demand that all interventions be evaluated by the same methodology. The US Government Performance Results Act of 1993 might be seen as such an example. Within certain domains that desire makes sense, and the tools of randomized-controlled clinical trials, evaluation science, and benefit-cost analysis all stand ready to implement such a call. Yet, when it comes to evaluating upstream interventions' effects on systems displaying significant delays and/or inertia, a more nuanced and context-specific approach to evaluation may be needed precisely because the before and after effects – even when a perfect counterfactual is available – do not look like a step function, immediate change in slope, or any other standard response of a system that does not have the same degree of inertia.

One final comment is in order. Evaluating interventions' effects on systems with nonlinear feedback is also difficult because effects depend not only on the input

(the intervention) but also on the system's state at the point of intervention. Systems with multiple equilibria separated by tipping points are a familiar example. The same intervention that has minimal effect if delivered when the system is near a stable equilibrium could have a dramatic effect if delivered at a moment such that it tips the system away from one equilibrium and toward another.

There is every reason to believe that at least drug systems exhibit such nonlinear feedback (Caulkins, 2005), and the same may be true of other domains such as violence as well. It is important to recognize, though, that none of the models or mathematics discussed above involves any such nonlinear feedback; they are all feed-forward and follow linear superposition. Nonlinearities exist and are a further complication, but "just" system lags and inertia alone are enough to make the contributions of upstream interventions difficult to appreciate. Since a large class of systems whose harms society seeks to prevent have such lags and inertia, observers who weigh in on the policy debate concerning the relative merits of upstream vs. downstream interventions need to factor this into their judgments.

## Appendix A: Justification for LH Model Parameter Values

The LH model has three parameters so one needs to know three facts about the system to pin down those parameters. For example, one might have at least a rough sense of what proportion,  $p$ , of people escalate to the H-state, the average “dwell time”,  $T$ , people spend in the H state if they reach it, and some notion of H individuals’ disproportionate participation. In drug and crime applications, typical values might be that  $p = 20\%$  of initiates become high-rate offenders who persist in that state for an average of  $T = 10$  or  $20$  years, collectively account for  $\pi = 80\%$  of the problem and, at an individual level, have offense rates that are  $k = 5$  or  $20$  times higher than those of low-rate offenders.

For the stylized model example in the body of the paper, we take  $T = 10$ ,  $p = 20\%$ ,  $\pi = 80\%$ , and  $k = 10$ . In an LH model, users spend an average of  $1/(a+b)$  years in the L state. The proportion  $p = b/(a+b)$  who reach the H state spend an average of  $1/g$  years there, so  $g = 1/T$  and an average of  $(b/g)/(a+b)$  years are spent in the H state per initiate, or  $b/g$  times longer than is spent in the L state.

Remembering that  $k$  is the ratio of H to L offense rates *per unit time*, that means  $\pi = \text{H offending} / (\text{L offending} + \text{H offending}) = (k b / g) / (1 + k b / g)$ , so  $b = (g / k) \pi / (1 - \pi)$ . Finally, since  $p = b/(a+b)$  of people reach the H state, we have  $a = b(1-p) / p$ . So for our stylized model these translate into  $a = 0.16$ ,  $b = 0.04$ , and  $g = 0.1$ .

To give specific examples that support this general tenor, in Greenwood et al.’s (1994) model of low- and high-rate criminal offenders, the  $p = 20\%$  who become high-rate offenders account for about  $\pi = 81\%$  of the crime despite their higher rates of incarceration because their individual offense rates are  $k = 17.52$  times higher and dwell times are twice as long ( $T = 20$  years) as for low-rate offenders.

For Australian injection drug users, Law et al. (2003) estimate an exit rate of 0.05 per year for dependent users (dwell time of  $T = 20$  years). Various authors (Weatherburn and Lind, 1995; Hall et al., 2000; Moore et al., 2005) assume that such users consume 17.5 times per week vs. once every week or two for recreational users, and the great bulk (perhaps  $\sim 70\%$ ) of all social costs are associated with the most problematic users (Caulkins et al., 2007).

Similar escalation probabilities ( $p = 0.16-0.17$ ; Anthony et al., 1994) and dwell times ( $T = 12 - 13$  years in Caulkins et al., 2004) have been estimated for US cocaine users. The best known studies use  $k$  values around 7 or 8. In particular,

Everingham and Rydell (1994) observed a  $k = 7.25:1$  rate among household survey respondents, and Abt (ONDCP, 2001) used similar ratios for weekly spending by chronic vs. occasional users of cocaine and heroin (averaging 7.4:1 and 8:1, respectively).

However, we believe higher  $k$  values are plausible. The Abt figures for more recent years assume spending by chronic users is in the vicinity of \$200 to \$250 per week, or \$10,000 - \$13,000 per year. That is consistent with various other reports. However, they assume weekly spending by occasional users averages \$35 per week. That might be realistic for a weekly user (roughly two \$20 purchases per week), but seems high as an average of all people who used less than once or twice per week, which is the cut-off Abt uses to distinguish recreational from chronic users.

The Household Survey is not helpful with respect to chronic cocaine users' spending, but it has some relevance for recreational users. In the 2005 Survey, those who used 52 times or fewer in the last year reported an average of 14.5 days of use in that year. If chronic users consume on an average of 300 days per year, the ratio  $k$  is 20.7 even if chronic users consume no more per day of use. Similar figures pertain for marijuana (14.3 times per year, so  $k = 21$  relative to those using 300 times per year), stimulants (14.9 times per year), and methamphetamine (16.1 times per year). Conditional on using no more than 52 times per year, the average days of use are well below 52, so the dollar value consumed per day would have to be quite high (over \$100) to bring  $k$  down to the 7 or 8 range.

For marijuana one might try to estimate  $k$  directly by comparing average days of use for those who used more than 52 times per year with the average days of use for those who used less than 52 times per year. The result is 214.8 vs. 14.3 days per year, suggesting  $k = 15$ . Parallel calculations yield  $k = 14.2$  for hallucinogens; for other drugs, the Household Survey misses too much heavy use for the calculations to be meaningful.

## References

- Anthony JC, Warner LA, Kessler RC (1994): Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. *Experimental and Clinical Psychopharmacology* 2:244-268.
- Behrens, Doris A., Jonathan P. Caulkins, Gernot Tragler, Josef Haunschmied, and Gustav Feichtinger. 1999. "A Dynamic Model of Drug Initiation: Implications for Treatment and Drug Control." *Mathematical Biosciences*. Vol. 159, pp.1-20.
- Behrens, Doris A., Jonathan P. Caulkins, Gernot Tragler, and Gustav Feichtinger. 2000. "Optimal Control of Drug Epidemics: Prevent and Treat – But Not at the Same Time." *Management Science*, Vol. 46, No. 3, pp.333-347.
- Blumstein, Alfred, Jacqueline Cohen, Jeffrey A. Roth, and Christy A. Visher, *Editors* (1986) *Criminal Careers and "Career Criminals"*, Volume 1. Panel on Research on Criminal Careers, Committee on Research on Law Enforcement and the Administration of Justice, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, DC.
- Caulkins, Jonathan P. 2005. "Models Pertaining to How Drug Policy Should Vary Over the Course of an Epidemic Cycle." In Bjorn Lindgren and Michael Grossman (eds.) *Substance Use: Individual Behavior, Social Interactions, Markets, and Politics*, Advances in Health Economics and Health Services Research, Volume 16. Elsevier, pp.407-439.
- Caulkins, Jonathan P. 2008. "Implications of Inertia for Assessing Drug Control Policy." *Contemporary Drug Problems*.
- Caulkins, Jonathan P., Doris A. Behrens, Claudia Knoll, Gernot Tragler, and Doris Zuba. 2004. "Modeling Dynamic Trajectories of Initiation and Demand: The Case of the US Cocaine Epidemic." *Health Care Management Science*. Vol. 7, No. 4, 319-329.
- Caulkins, Jonathan P., Paul Dietze, and Alison Ritter. 2007. "Dynamic Compartmental Model of Trends in Australian Drug Use." *Healthcare Management Science*. Vol 10., No. 2. DOI: 10.1007/s10729-007-9012-0.
- Chen, K. and D.B. Kandel (1995) The natural history of drug use from adolescence to the mid-thirties in a general population sample. *American Journal of Public Health*. 85(1):41-47.
- Crane, B.D., A.R. Rivolo, and G.C. Comfort. 1997. *An Empirical Examination of Counterdrug Interdiction Program Effectiveness*. Alexandria, Virginia: Institute for Defense Analysis.
- Everingham, S.S. and C.P. Rydell: (1994). *Modeling the Demand for Cocaine*. MR-332-ONDCP/A/DPRC, RAND: Santa Monica, CA.

- Farrington, D.P. (2003) Developmental and Life-Course Criminology: Key Theoretical and Empirical Issues – The 2002 Sutherland Award Address. *Criminology*. 2:221-256.
- Graham, Hilary, Hazel M Inskip, Brian Francis and Juliet Harman (2006) Pathways of disadvantage and smoking careers: evidence and policy implications. *J. Epidemiol. Community Health*;60;7-12, doi:10.1136/jech.2005.045583.
- Greenwood, Peter W., C. Peter Rydell, Allan F. Abrahamse, Jonathan P. Caulkins, James Chiesa, Karyn E. Model, and Stephen P. Klein. (1994). *Three Strikes and You're Out: Estimated Benefits and Costs of California's New Mandatory Sentencing Law*, MR-509-RC, RAND, Santa Monica, CA.
- Hall, W., Ross, J., Lynskey, M., Law, M. & Degenhardt, L. (2000). How many dependent opioid users are there in Australia?, NDARC Monograph No. 44. Sydney: National Drug and Alcohol Research Centre.
- Hser, Yih-Ing, M. Douglas Anglin, Christine Grella, Douglas Longshore, and Michael L. Prendergast. 1997. Drug Treatment Careers: A Conceptual Framework and Existing Research Findings. *Journal of Substance Abuse Treatment*. 14(6), 543-558.
- Hser, Yih-Ing, Douglas Longshore, and M. Douglas Anglin. 2007a. The Life Course Perspective on Drug Use; A Conceptual Framework for Understanding Drug Use Trajectories. *Evaluation Review*. 31(6), 515-547.
- Hser, Yih-Ing, David Huang, Chih-Ping Chou, and M. Douglas Anglin. 2007b. Trajectories of Heroin Addiction: Growth Mixture Modeling Results Based on a 33-Year Follow-up Study. *Evaluation Review*. 31(6), 548-562.
- Johnson, Bruce D., Andrew Golub, and Jeffrey Fagan. 1995. "Careers in Crack, Drug Use, Drug Distribution and Nondrug Criminality." *Crime and Delinquency* 41(3):275-295.
- Law, M.G., G.J. Dore, N. Bath, S. Thompson, N. Crofts, K. Dolan, W. Giles, P. Gow, J. Kaldor, S. Loveday, E. Powell, J. Spencer, and A. Wodak. (2003). Modelling hepatitis C virus incidence, prevalence and long-term sequelae in Australia, 2001. *International Journal of Epidemiology* 32, 717-724.
- Makkai, Toni and Jason Payne (2003) Key findings from the drug use careers of offenders (DUCO) study, Australian Institute of Criminology, Canberra, Australia.
- Moore, T.J., J.P. Caulkins, A. Ritter, P. Dietze, S. Monagle, and J. Pruden, Heroin Markets in Australia: Current Understanding and Future Possibilities. DPMP Monograph Series. (Turning Point Alcohol and Drug Centre, Fitzroy, 2005).
- Nagin, D. S. 2005. *Group-based Modeling of Development*. Cambridge, MA.: Harvard University Press.

Nagin, D.S., and Tremblay, R.E. 2005. "What Has Been Learned from Group-based Trajectory Modeling?: Examples from Physical Aggression and Other Problem Behaviors." *Annals of the American Academy of Political and Social Science*, 602: 82-117.

Office of Applied Studies (OAS). (2002). *National and State Estimates of the Drug Abuse Treatment Gap: 2000 National Household Survey on Drug Abuse* (NHSDA Series H-14, DHHS Publication No. SMA 02-3640). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of National Drug Control Policy (ONDCP) 2001. *What America's Users Spend on Illegal Drugs*. Washington, DC: The White House.

Office of National Drug Control Policy (ONDCP) 2004. *The Economic Costs of Drug Abuse in the United States, 1992-2002*. Washington, DC: The White House.

Sampson, R.J. and J.H. Laub (2005) A Life-Course View of the Development of Crime. *Annals of the American Academy of Political and Social Sciences*. 602(1):12-45.

Substance Abuse and Mental Health Services Administration, Office of Applied Studies. Substance Dependence, Abuse and Treatment: Findings from the 2000 National Household Survey on Drug Abuse. OAS Analytic Series #A-16, DHHS Publication No. (SMA) 02-3642, Rockville, MD, 2002.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2006). *Results from the 2005 National Survey on Drug Use and Health: National Findings* (Office of Applied Studies, NSDUH Series H-30, DHHS Publication No. SMA 06-4194). Rockville, MD.

Weatherburn, D., & Lind, B. (1995). *Drug law enforcement policy and its impact on the heroin market*. Sydney: NSW Bureau of Crime Statistics and Research.

Weatherburn, Don, Craig Jones, Karen Freeman, and Toni Makkai. (2003) Supply Control and Harm Reduction: Lessons from the Australian Heroin 'Drought'." *Addiction*, Vol. 98, pp.83-91.

Wright, D. (2004). *State Estimates of Substance Use from the 2002 National Survey on Drug Use and Health* (DHHS Publication No. SMA 04-3907, NSDUH Series H-23). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.