Treatment on demand: an operational model [*]

Edward H. Kaplan [a] and Mira Johri [b]

[a] *Yale School of Management, Box 208200, New Haven, CT 06520-8200, USA*
E-mail: edward.kaplan@yale.edu
[b] *Yale School of Medicine, Department of Epidemiology and Public Health, 60 College St., Box 208034, New Haven, CT 06520-8034, USA*

Received 24 June 1999; accepted 29 October 1999

The goal of "treatment on demand" is to allow all those seeking substance abuse treatment immediate entry into a program. Surprisingly, little is understood regarding the relationship between the demand for treatment, queues, waiting times and treatment admission rates, and treatment capacity. Nor has the increase in treatment capacity required to eliminate drug treatment queues, along with the expected benefits and costs of such an expansion, been studied carefully. In this paper, we present a mathematical model of drug treatment flows for systems where the demand for treatment greatly exceeds available supply. The model produces estimates of queue lengths, waiting times and treatment admission probabilities for any given treatment capacity, and suggests the capacity needed to achieve treatment on demand. The model also enables one to contrast the likely costs and benefits of changes in treatment capacity. We illustrate the model using San Francisco as a case study.

1. Introduction

"Treatment on demand" is an often-mentioned phrase in the contemporary drug policy debate. The goal is to allow all those seeking substance abuse treatment immediate entry into a program by augmenting system capacity to a sufficient level. The arguments for treatment on demand follow from two claims. The first regards the benefits of drug treatment in countering the human and economic costs of drug abuse [1–4], while the second is the reported shortage of available drug treatment slots [5–9]. Indeed, in the United States, the cities of Baltimore and San Francisco have officially committed themselves to providing drug treatment on demand, and other communities are poised to do so [7,10].

Clearly, providing treatment on demand requires increasing the number of available drug treatment slots. Surprisingly, however, very little is known regarding the relationships between treatment request rates, queue lengths, waiting times and treatment admission probabilities, and the capacity of drug treatment programs. To our knowledge, no one has indicated the increases in treatment capacity that treatment on demand would require, let alone estimated the attendant costs and benefits of such a strategy.

This paper makes an initial attempt to answer these questions via a mathematical model of drug treatment flows in systems where the demand for treatment greatly exceeds available supply. The model produces estimates of queue lengths, waiting times and treatment admission probabilities for any given treatment capacity over time. The model suggests the capacity needed to achieve treatment on demand, as well as the time frame required to eliminate treatment queues. The model also enables one to contrast costs and benefits of changes in treatment capacity.

In the next section, we discuss several operational issues surrounding treatment on demand that our model can help to resolve. Following this, we introduce the assumptions that underlie the model, and consider how departures from these assumptions would affect the model. We illustrate the model using San Francisco as a case study in section 4. A mathematical description of the model appears in the appendix.

2. Treatment on demand: operational issues

2.1. Treatment queues, waiting times and service levels

Whether treatment on demand is an achievable goal depends squarely upon the operational terms that are used to define it, and there are alternatives to consider. If "immediate entry to treatment" corresponds to "no waiting" under ideal circumstances, then treatment on demand should imply the elimination of drug treatment queues (for a treatment queue is the collection of those waiting for treatment). And, if the goal is to admit all those seeking treatment, then treatment admission probabilities (or "service levels", defined as the fraction of those requesting treatment who actually receive treatment) should equal 100%.

However, pragmatism demands at least a minor retreat from such ideal standards in practice. For example, the stated goal of treatment on demand in San Francisco is placement within 48 hours of a treatment request [10]. While reasonable, such a definition ignores the impact of

[*] Research supported in part by the Societal Institute for the Mathematical Sciences via grant DA-09531 from the National Institute on Drug Abuse, and the Center for Interdisciplinary Research on AIDS at Yale University via grant PO1-MH/DA-56826 from the National Institutes on Mental Health and Drug Abuse. We also acknowledge the participants in the SIMS Policy Modeling Meeting, January 12–13, 1998, Washington, DC (Dr. James G. Kahn, Chair), and Celia Ponvert for research assistance.

inserting delay on queues and service levels. For example, our model suggests (see equation A.10 in the appendix) that depending upon the length of time drug users are willing or able to wait for treatment, a 48 hour waiting time could be consistent with high service levels (94% if those requesting treatment can wait an average of 1 month before dropping out), moderate service levels (75% if persons can wait an average of 1 week), or low service levels (14% if persons can wait only 1 day on average). Although we are not aware of any formal studies of the time users are willing (or able) to wait before dropping out, observational reports by those involved in drug treatment suggest that it is likely to be considerably less than one month, and perhaps only a matter of days. One expert reports that at Beth Israel Hospital in New York, "Our experience is that after several months on a waiting list, about half the prospective patients wind up lost to contact. We have also asked research storefront recruited subjects whether they are interested in entering treatment, and about 80% say they are if they could get into treatment that day" [11]. What seems like a reasonable implementation of treatment on demand could thus end up admitting fewer than 15% of all those requesting treatment!

Indeed, any particular policy governing the capacity of the treatment system implies consequences that can be measured in various ways. We have chosen to focus on treatment queues, waiting times and service levels as operational measures that are particularly important when considering drug treatment on demand.

2.2. Capacity required for treatment on demand

How many treatment slots are required to achieve treatment on demand? In the short run, eliminating the treatment queue would require adding one slot for each drug user currently waiting for treatment. Maintaining immediate access to treatment could well require even more slots to accommodate additional treatment requests in the short run. However, the number of incremental slots required for treatment on demand over the long run could be greater or less than the existing queue for treatment. For example, our model suggests that if the average time users are willing (or able) to remain in the treatment queue exceeds the average duration of a drug treatment episode, then the number of incremental slots required is less than the length of the existing queue. Our model produces estimates of the number of treatment slots required to eliminate treatment queues over time. The model can also estimate the capacity required for weaker definitions of treatment on demand (such as provision of treatment within 48 hours of request).

2.3. The direct and indirect benefits of treatment on demand

The benefits of increasing the number of drug treatment slots are realized as an increasing number of persons are removed from active drug use. To the extent that important variables of societal concern such as crime rates, employment, the incidence of infectious diseases such as HIV, and child abuse are different for those who use drugs and those who do not, the impact of changing drug treatment capacity can be expected to extend to such variables via changes in the number of active drug users over time [1,2].

When examining the benefits of drug treatment, there are two different effects to consider. The *direct* benefits of drug treatment follow from the utilization of a given slot, for while an individual is in treatment, presumably drug use is reduced if not eliminated. The direct benefits of increasing the number of treatment slots are therefore realized as soon as the additional slots are filled. Even if all treated individuals relapse to drug use immediately following treatment, these direct benefits would still be realized. The *indirect* benefits of drug treatment occur when persons completing drug treatment remain drug free for appreciable lengths of time. These benefits accrue over time, and depend critically upon both the fraction of treatment episodes that are successful in eliminating (or greatly reducing) drug use, and the time until relapse to drug use following drug treatment. Our model helps illuminate both the magnitude of these benefits (as measured in incremental drug free years per slot added), as well as the timing of the indirect effects. The model also reminds us that there is a limit to the overall benefits that treatment on demand can provide: only those drug users seeking (or remanded to) treatment can be helped. It is therefore unreasonable to presume that treatment on demand can end the drug problem. Indeed, we do not even argue here that treatment on demand provides the most cost-effective path towards any particular target for the aggregate level of substance abuse relative to other possible policies, for analysis of competing policies (such as education and prevention or enforcement) is beyond the scope of this paper [3].

3. Modeling drug treatment flows: assumptions

In the discussion that follows, we do not distinguish between types of drugs, treatment modalities, or different client demographics. Our aim is to model general relationships that hold at a high level of abstraction, with the idea that these same principles can be adopted for more specialized study of specific treatment populations and programs. We will now detail the assumptions underlying our model, and discuss the utility such a model provides for understanding the relationships between drug treatment capacity, operational variables such as treatment queue lengths, service levels and waiting times to enter programs, and the benefits that accrue from changing the capacity of the drug treatment system.

Consider a fixed population of drug users (closed to arrivals and departures, a reasonable assumption over short time frames). The model assumes that at any moment in time, a drug user can be in one of four states of behavior (figure 1): abstinent; using drugs but *not* waiting for drug

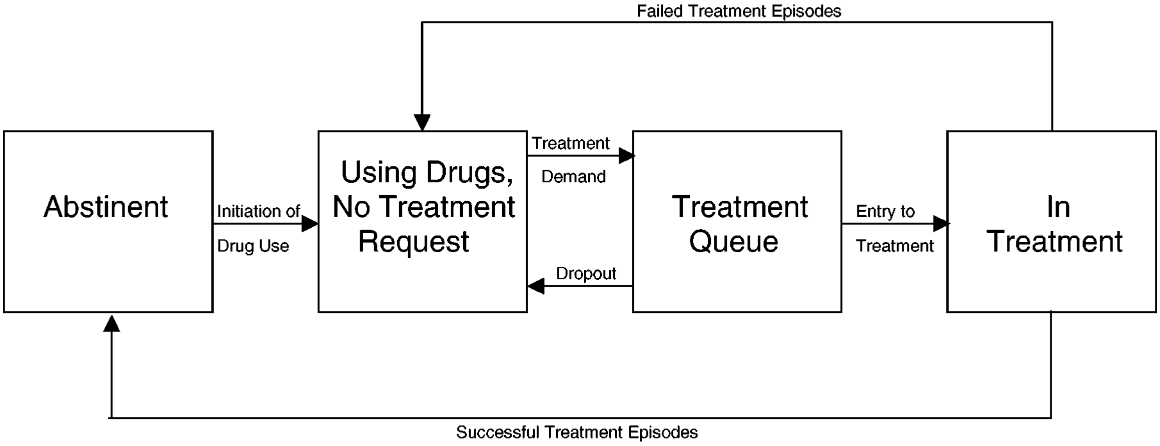

Figure 1. A model of drug treatment flows.

treatment; using drugs and waiting for treatment; or in drug treatment (and, by assumption, not using drugs). With such a model, it is possible to estimate population flows between states of behavior over time, resulting in forecasts of the number of drug users in each state of behavior at future times, from more basic postulates governing the behavior of drug users.

Population flows between states are assumed to proceed as follows. All abstinent individuals eventually initiate (or return to) drug use without having requested treatment. In the model, all (temporarily) abstinent users initiate (or relapse into) drug use at the same rate, so the aggregate flow from the abstinent to the drug use absent treatment request state equals the product of the number of abstinent users and a drug use initiation (or relapse) rate. We recognize that empirically, relapse rates depend on the length of time since the completion of treatment, and as such are not constant over time [1,12–14]. However, since we are concerned with the *aggregate* relapse rate rather than the recidivism of any specific individual, the constant relapse rate assumption will not greatly affect our results.

We assume that persons using drugs while not waiting for treatment request entry (or are remanded) to drug treatment at a constant per capita rate. An objection to this assumption might be that individual demand for treatment could depend upon the amount of treatment available. Increasing the number of treatment slots could lead to an increase in the demand for treatment by this argument. Since our model does not allow such demand responsiveness, it may well underestimate the demand for treatment as a function of the number of offered slots. This translates into an optimistic bias when we estimate the capacity required to ensure treatment on demand, in that the actual number of slots required would be *at least* as large as what our model suggests.

Because demand exceeds available treatment capacity, individuals are forced to wait in a queue for treatment. Once in queue, one of two events takes place. First, a waiting individual may become discouraged by the waiting process (or find himself removed from the queue due to committing a crime or parole violation) and drop out of the treatment queue. The aggregate dropout rate is modeled as proportional to the number of persons waiting in queue, implying a constant dropout rate for those waiting. Dropouts are assumed to return to the state of drug use absent a request for (or remand to) treatment. Second, individuals who do not drop out of the queue proceed to treatment. Assuming a saturated system, the aggregate rate with which individuals enter treatment programs equals the aggregate departure rate from treatment (i.e., the turnover rate). We assume that the average duration of treatment is independent of the number of treatment slots available.

In keeping with evidence that drug use is a chronic condition marked by cycles of use, dependency and recovery, the consequences of treatment are depicted in a deliberately conservative fashion. Once in treatment, one of two outcomes must occur: individuals either successfully complete treatment, or fail. Treatment successes return to the group of abstinent users, while treatment failures are assumed to return directly to drug use without waiting for treatment. The probability that a treatment episode successfully returns an individual to the abstinent state is assumed independent of the number of slots available. This assumption could be questioned, for it may well be that those who currently survive treatment queues and are admitted to treatment are more likely to succeed than those who drop out; if so, then increasing the number of slots would *reduce* the probability of a treatment episode reaching a successful conclusion. The bias introduced by this assumption, however, also serves to understate the capacity required to provide treatment on demand. Again, the treatment capacity needed in truth will be *at least* as large as what the model prescribes.

Formalizing the treatment flows described above (and illustrated in figure 1) into the mathematical model detailed in the appendix suggests answers to several questions, including: As a function of the number of drug treatment slots, how many drug users will be waiting to enter drug treatment at future times? How many slots are needed to eliminate treatment queues? How long will it take to eliminate treatment queues? How long must drug users requesting treatment wait to gain entry to a program? What fraction of treatment requests actually gain entry to treatment? In addition, the model suggests how the population will redistribute between those using drugs (whether waiting for treatment or not) and those not using drugs (whether in treatment or not). This latter calculation enables the estimation of the change in drug free years across the population as a function of the number of drug treatment slots made available.

4. Modeling drug treatment on demand: an illustration

We will now illustrate the model using San Francisco, California as a case study, emphasizing the input data employed and the results obtained. Sufficient references to the appendix are provided to enable readers to reproduce the analysis reported below.

Alcohol and other drug-related problems have reached critical levels in San Francisco. The city's Department of Public Health estimates that the local cost of drug abuse and related problems was $1.7 billion in 1996 [10]. In light of the severity of the problem, the Board of Supervisors resolved in November of 1996 to support the goal of substance abuse treatment on demand for residents of San Francisco [15]. It endorsed full funding for implementation of a program of expanded, relevant and prompt treatment for the city's substance abusers. Policy makers for San Francisco have defined a target of admission to treatment within two days [10].

San Francisco provides an excellent test case for our model, for two reasons. First, the city clearly meets a critical assumption of the model; namely, that the demand for treatment greatly exceeds supply. Second, San Francisco's political commitment to increase its spending on drug treatment has led to the collection of relevant data, such as the number of drug users and the length of the treatment queue, that are missing in other municipalities and at the national level [16].

4.1. Model inputs

We specified the model parameters using a combination of US and San Francisco data. National figures were used for the average duration of treatment, percentage of participants completing treatment, and average time spent drug free following treatment episodes. All other parameter values were taken from the Department of Public Health for San Francisco, or estimated via the model assuming that the current treatment system in San Francisco had reached equilibrium (see section A.4 in the appendix). The model inputs and their sources are summarized in table 1.

4.2. Treatment queues, waiting times and service levels

San Francisco's Director of Public Health reports an estimated 45,000 drug users in San Francisco [17]. Of these, 1,400 are reported to be waiting to enter one of the approximately 6,300 treatment slots available [17]. Following figure 1, of the remaining 37,300 persons in the population of

Table 1
Parameter estimates.

Parameter	Notation (see appendix)	Estimate	Source
Population of drug users in need of treatment	n	45,000	[17]
Average number of persons waiting in queue for drug treatment	q	1,400	[17]
Average number of persons abstinent	a	17,618	Consistency requirement assuming steady state; see section A.4 of the appendix for details
Number of drug treatment slots	s	6,328	[17]
Annual turnover rate per drug treatment slot	μ	3.5	Midpoint value, estimated as the reciprocal of the average duration of treatment (which falls between 0.25 and 0.34 years [3,20,21])
Probability of success per treatment episode	p	0.175	[1]
Yearly drug use initiation rate from abstinence	ι	0.22	[2]
Annual demand rate for treatment	α	$1.127 + 0.071\delta$	Forces consistency between the model and the reported queue length of 1,400 in San Francisco; see section A.4 of the appendix for details.
Tolerance for delay: the average time a user is willing and/or able to wait before dropping out of the treatment queue	$1/\delta$	Unknown: scenarios studied setting the tolerance for delay to one year, one month, one week, and one day	

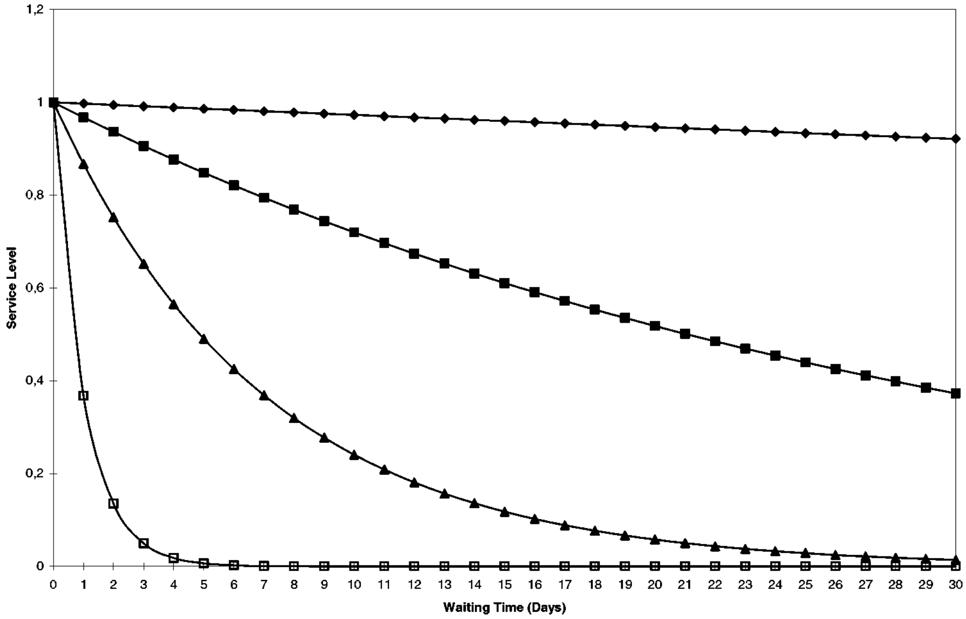

Figure 2. Service levels and waiting times for delay tolerances ($1/\delta$) of 1 year (♦), 1 month (■), 1 week (▲) and 1 day (□).

Table 2
Waiting times and service levels.

Tolerance for delay ($1/\delta$)	Waiting time (days)	Service level (%)
1 year	22	94
1 month	17	57
1 week	10	23
1 day	3	4

drug users, we estimate (via equation (A.27) in the appendix) that about 17,600 users are currently abstinent, leaving some 19,700 active drug users who have neither requested nor been remanded to treatment programs. The model thus suggests that of the 21,100 *active* drug users in the population, only 7% are waiting for treatment.

Lacking data reporting the fraction of treatment requests that actually receive drug treatment or the typical waiting time to treatment for those who do obtain service, we turn to our model. As argued earlier, the waiting times and service levels depend critically upon the length of time drug users are willing (or able) to wait for treatment. We consider four possible values of this *tolerance for delay*: users on average can wait 1 year after requesting treatment before dropping out of the treatment queue; or alternatively 1 month; 1 week; or 1 day.

Via equations (A.8) and (A.10) in the appendix, we report the waiting times and service levels that are consistent with an observed treatment queue of 1,400 drug users in table 2. If one accepts the common belief that users are not able to endure long waiting periods in treatment queues, then the waiting times to enter drug treatment in San Francisco are estimated to be on the order of a few days to a week or two – for those who get in. Unfortunately, table 2 also reveals that less than half of those requesting treatment will actually receive service if the tolerance for delay is short. Indeed, even if applicants remain committed to entering treatment for an average of one month following their request, only 57% would survive the wait associated with the existing 1,400 person treatment queue.

We do not intend to mislead the reader: for a *fixed* tolerance for delay, shorter waiting times do imply higher service levels, as illustrated in figure 2 (which is also based on equation (A.10)). Even so, note that small increases in waiting times can imply large reductions in the service level. Setting what seems like a reasonable goal for waiting times could result in unacceptably low admission rates to treatment programs.

A reviewer of this article noted that the reported queue length is likely to be an overestimate of the actual number of people waiting to enter treatment on account of duplicate record keeping and failure to purge individuals no longer waiting for treatment. (S)he therefore asked how our results would change if the initial queue length encountered was equal to 1,000 or 500 persons (as opposed to 1,400). At one extreme, if the tolerance for delay is as long as one year, then the average waiting time would fall from 22 days (when the queue contains 1,400 people) to 16 and 8 days (for initial queues of 1,000 and 500, respectively), while service levels for the corresponding queue lengths would rise from 94% to 96% and 98%. At the other extreme, if the tolerance for delay is as short as one day, then the mean waiting times would fall from 3 days to 2 days while service levels would increase from 4% through 11% (as the initial queue falls from 1,400 to 500). It should not be surprising that for any given delay tolerance, shorter queues coincide with shorter waiting times and higher service levels as mentioned previously.

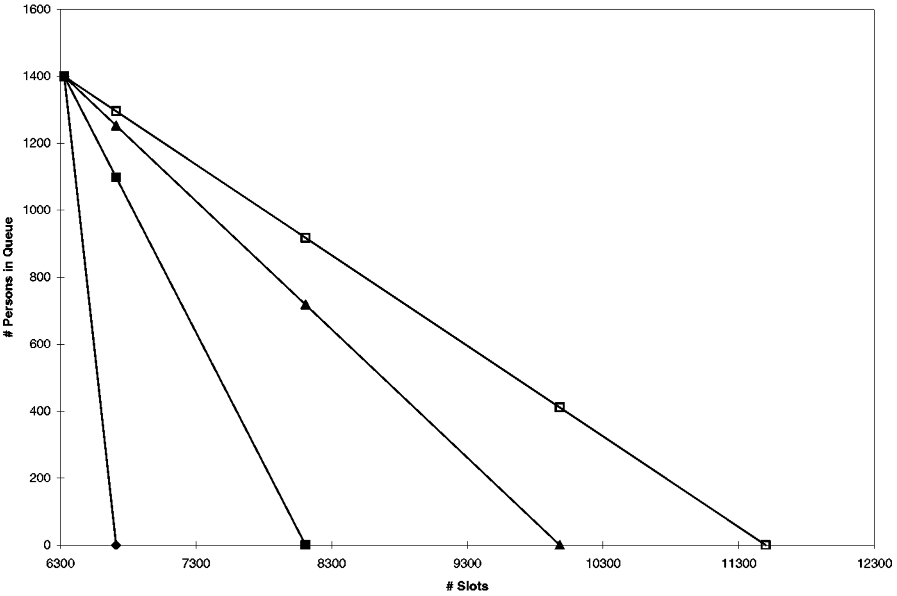

Figure 3. Long run queue lengths as a function of treatment capacity for delay tolerances ($1/\delta$) of 1 year (♦), 1 month (■), 1 week (▲) and 1 day (□).

4.3. Capacity required for treatment on demand

4.3.1. The long run

In San Francisco, how many slots are needed to eliminate treatment queues in the long run? Following section A.2.1 in the appendix and assuming the parameter values of table 1, figure 3 reports steady state treatment queue lengths as a function of the number of available slots (equation (A.11) in the appendix). Recall that a naïve guess for the capacity required to eliminate the treatment queue is "existing number of slots" plus "queue length", or $6,328 + 1,400 = 7,728$. Figure 3 reveals significant departures from this naïve estimate. If the tolerance for delay averages one year, the number of slots needed is only 6,710, which is appreciably less than the naïve guess. However, for the three other scenarios, the number needed is higher, sometimes considerably higher. When average willingness or ability to wait is one month, the required number of slots equals 8,110; for one week, 9,980; and for one day, 11,500. These figures translate into substantial differences in expected costs. The approximate average cost of one drug treatment slot in San Francisco is $6,000.[1] Using the naïve guess, one would budget approximately $1,400 \times 6,000 = \$8,400,000$ in additional funds. From figure 3, however, the required incremental costs could run as low as $2.3 million, or as high as $31 million dollars.

How would these results change if the existing queue for treatment in San Francisco was equal to 1,000 or 500 instead of the 1,400 reported in table 1? Perhaps surprisingly, the results would not change drastically at all. For delay tolerances ranging from one year to one day, the capacity required to establish treatment on demand consistent with an initial queue of size 1,000 is reduced to 6,600, 7,700, 9,500 and 11,350 slots, respectively. If the initial queue was instead equal to 500, the required numbers of slots would fall further to 6,500, 7,100, 8,500 and 10,900 over the same mean tolerances for delay. For a given tolerance of delay, there are no consequential differences in the number of slots required to achieve treatment on demand on account of uncertainty in the initial size of the treatment queue in San Francisco.

What explains these results? The issue is one of calibration. To maintain consistency with the estimated distribution of drug users in San Francisco over the four behavioral states of the model (as shown in figure 1), the estimated demand rate for treatment must shrink as the initial queue length is reduced (see equation (A.32) in the appendix), explaining why the number of treatment slots needed to eliminate waiting falls with the initial queue length. However, consistency with the observed queue length also requires that the estimated demand rate for treatment must grow as the tolerance for delay is reduced (again see equation (A.32) in the appendix). Thus, as the tolerance for delay is reduced in the examples above, the demand for treatment *increases* in response, increasing the number of slots required to eliminate all waiting. The impact of changing the tolerance for delay is clearly much greater than the impact of initial queue lengths in the examples considered above.

Recall that in San Francisco, treatment on demand is defined as providing treatment within 48 hours. Figure 4 reports the long run number of slots required to provide a family of "weak" treatment on demand policies defined by maximum waiting times (following equation (A.16) in the appendix). To ensure waiting times of two days, for example, figure 4 reports capacity requirements of 6,680, 7,890, 9,410 and 9,350 slots for delay tolerances running from 1 year through 1 day, respectively. Again, defining

[1] This was calculated as follows: the current budget for substance abuse programs in San Francisco is $37.8 million [17]. If there are 6,328 slots, this works out to a cost of $5,973.45 per slot. Note that this represents a combination of inpatient and outpatient treatment. A residential treatment slot costs approximately $21,000 annually [8].

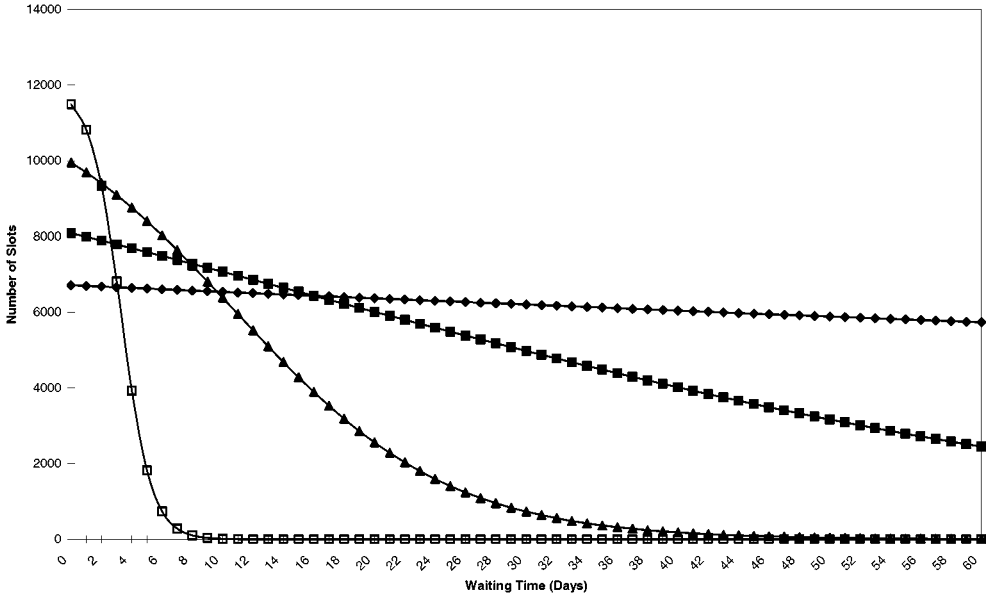

Figure 4. Long run capacity required for "weak" treatment on demand for delay tolerances ($1/\delta$) of 1 year (♦), 1 month (■), 1 week (▲) and 1 day (□).

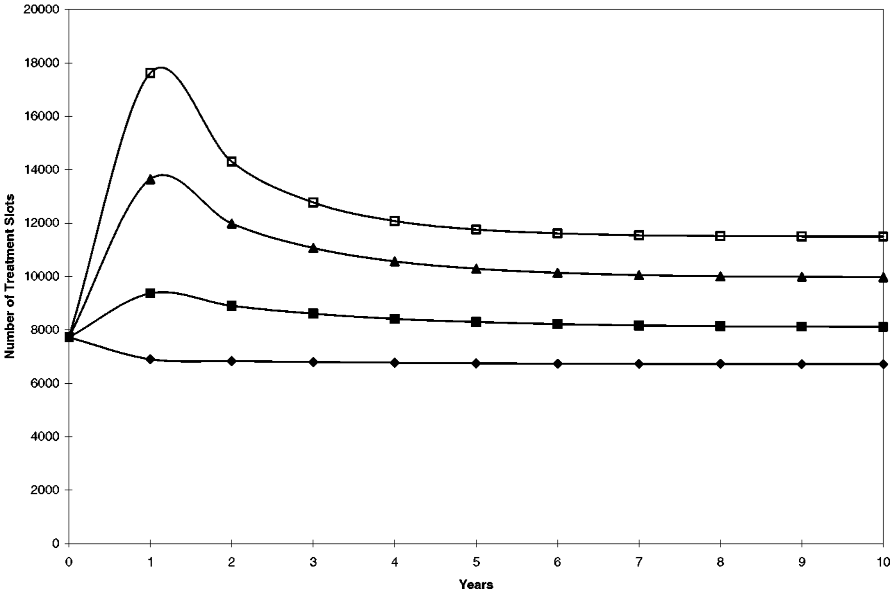

Figure 5. Capacity required to eliminate treatment queue for delay tolerances ($1/\delta$) of 1 year (♦), 1 month (■), 1 week (▲) and 1 day (□).

treatment on demand in this fashion does not reveal the entire picture, for the service levels associated with these scenarios differ substantially depending upon the tolerance for delay, ranging from 99% (1 year tolerance) to 14% (1 day tolerance).

4.3.2. *The short run*

Increasing the current number of slots to the steady state levels required for treatment on demand by whatever definition will not result in immediate achievement of long run results. Immediate results require more drastic action. Figure 5, produced via equations (A.20)–(A.24) in the appendix, reports the number of treatment slots required to eliminate treatment queues immediately and over time. Initially, 1,400 new slots must be added to the 6,300 already in existence to instantaneously erase the treatment queue. Following this, if the tolerance for delay is relatively short (1 day or 1 week), the required number of slots must grow rapidly to nearly 18,000 slots (over 50% of the steady state level for a 1 day tolerance for delay), and to 14,000 slots (over 40% of the steady state level for a 1 week tolerance of delay). This effect is less pronounced for a 1 month tolerance for delay, while if applicants actually were able to spend an average of one year waiting for treatment before dropping out, the number of slots required over time would actually decline.

It is important to remind the reader that it is *not* the tolerance for delay *per se* that is driving these results. Rather, it is the differences in the demand for treatment *implied* by different tolerances for delay combined with the observed treatment queue in San Francisco. The scenarios above that require more treatment slots are precisely those scenarios with greater demands for treatment (as explained in section A.4 in the appendix). This makes figure 5 easier to understand: the greater the demand for treatment, the greater the number of treatment slots required to provide treatment on demand.

4.4. The direct and indirect benefits of increasing treatment capacity

The benefits of increasing the capacity of drug treatment derive from increasing the total number of drug free years in the population of drug users. To characterize the nature and timing of such benefits, we assume that both persons in treatment and those abstinent are not currently using drugs, while all others in the population are using drugs. In the long run, the model suggests that as long as the total number of slots remains below the capacity required to eliminate the treatment queue, each incremental treatment slot created will produce an additional 3.78 drug free years annually (as derived via equation (A.29)). If the number of slots exceeds the capacity required to eliminate the queue in steady state, then no benefits are derived (and resources are being wasted).

That benefits accrue at the rate of 3.78 drug free years per slot added in steady state can be understood as follows: assuming that the current number of slots is less than the critical number needed to zero the queue (and hence that additional slots are always filled), the direct effect of adding a treatment slot will be to produce one drug free person year. The indirect effect of adding a treatment slot is due to the benefits accruing to successful completion of treatment. From table 1, we note that since (i) each slot turns over an average of 3.5 times per year; (ii) 17.5% of treatment episodes lead to successful treatment completions; and (iii) each success buys 4.5 drug free years on average, the expected indirect benefit per slot is $(3.5 \times 0.175 \times 4.5)$, or 2.78 drug free years annually. The total benefit per incremental slot measured in terms of drug free years is hence the sum of direct and indirect benefits, or $1 + 2.78 = 3.78$ drug free years.

We remind the reader at this point that if one believes that the average effectiveness of treatment episodes declines as the number of slots increases (for example, due to self-selection of drug users committed to completing treatment), then this calculation would change, as the indirect effect would be lower than the 2.78 years estimated above. The direct effect would not change, however. Thus, in addition to lower bounds for the capacity required for treatment on demand, our model produces upper bounds for the benefits derived from increasing treatment capacity.

The discussion above addressed the long run benefits of treatment. In the short run, however, things are different. Figure 6 reports the average increase in drug free years per additional treatment slot over time (via equation (A.31) in the appendix). In the very short run, only 1 drug free year is gained per treatment slot. This of course is exactly the direct effect of treatment. Over time, however, the benefits rise to 3.78 years. The time delay is exactly the time required for new treatment successes (and their associated drug free years) to propagate through the treatment system. Figure 6 thus shows that the indirect effect, though ultimately substantial, is a significantly lagged benefit. And, if the mean effectiveness of treatment declines as the number of slots increases, the indirect effect would be diminished even further.

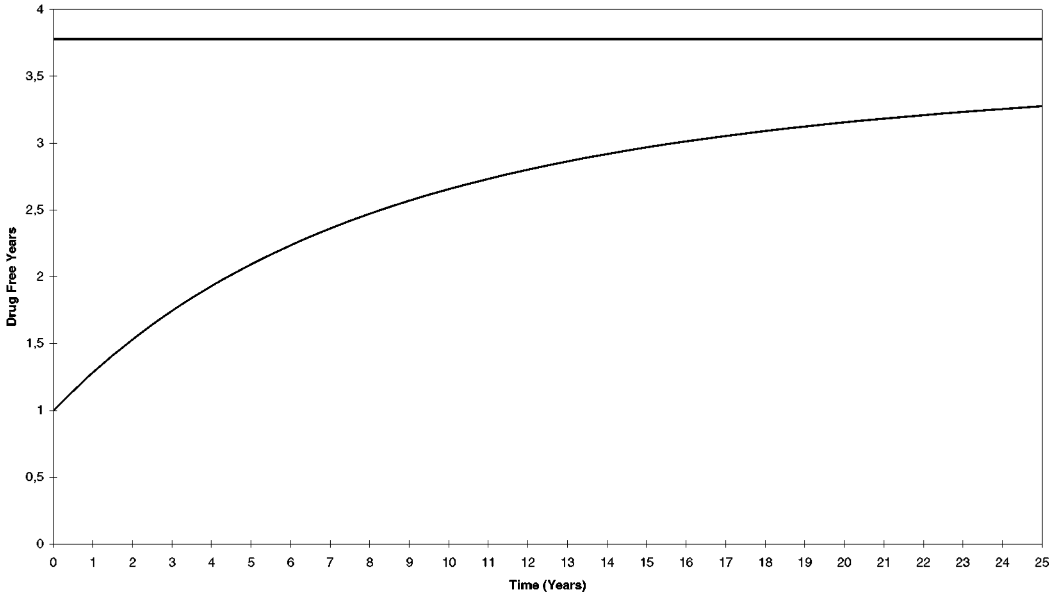

Figure 6. Average annual incremental drug free years per treatment slot.

5. Discussion

Although treatment on demand is prominent in contemporary drug policy, we possess little understanding of what this goal would entail in terms of increases from existing capacity and expected benefits. The model of this paper makes a unique contribution to the debate by proposing an approach to rigorously study these questions. Sample results based in part on data from San Francisco indicate that the actual number of slots needed to zero the treatment queue may well be greater than indicated by inspection of queue lengths, and that the time frame involved is considerable. Retreating to weaker definitions of treatment on demand (such as ensuring placement within 48 hours) carries the serious risk of losing many if not most of those who request treatment before admission to a program can be offered. Nonetheless, our analysis joins a growing number of studies suggesting that there are important benefits to drug treatment [3,4,18]. Figure 6 shows that additional treatment slots could return 2 drug free years per slot annually within 5 years, while we estimated that in the long run, this benefit would rise to 3.78 drug free years per slot per year.

Data limitations combined with a desire to maintain mathematical tractability led us to make certain simplifying assumptions that impact the analysis in various ways worth repeating. We do believe that on account of simplifying assumptions, our model is likely to understate the true number of slots required to eliminate treatment queues, and overstate the marginal benefit delivered per additional treatment slot. Worded differently, having employed this model, a reasonable interpretation would be that conditional on the model inputs employed, the elimination of all waiting for drug treatment programs requires *at least* the number of slots we have prescribed in the various scenarios, while the benefits obtained from doing so (measured in drug free years) will *at most* equate to the figures stated.

We have constructed a model at a high level of aggregation and abstraction. Clearly more serious study could be directed at specific drugs, treatment programs, and user populations. Careful monitoring of treatment requests, queue composition, the duration of treatment episodes, and treatment outcomes in a specific program could lead to much better estimates of quantities such as the demand for treatment, the tolerance for delay, and treatment turnover and success rates. Indeed, improved record keeping and treatment queue management are required for any serious empirical study of the consequences of changing drug treatment capacity.

Nonetheless we stand by the key relationships the model has uncovered, such as the interplay between queues, waiting times and service levels, and treatment capacity, and the decomposition of treatment benefits into direct and indirect effects. The numerical values will of course be different in different circumstances, but the basic relationships identified are, we believe, fundamental.

Appendix. Mathematical model of drug treatment flows

As explained in the text, the model divides the (presumed fixed) population of drug users into four compartments (or *states*). Let:

$n =$ the number of drug users in the population,

$a(t) =$ the number in the population who are abstinent from drug use at time t,

$q(t) =$ the number using drugs but waiting for treatment (i.e., in queue) at time t,

$s =$ the (constant) number of users in treatment at time t, and

$n - s - q(t) - a(t) =$ the number using drugs and not waiting for treatment at time t (since the total population is fixed and equal to n).

The dynamics of this model can be represented mathematically as:

$$\frac{\mathrm{d}a(t)}{\mathrm{d}t} = s\mu p - a(t)\iota \tag{A.1}$$

and

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = \bigl(n - s - q(t) - a(t)\bigr)\alpha - q(t)\delta - s\mu \tag{A.2}$$

where:

$\iota =$ the annual per capita “initiation rate” from abstinence to drug use (so the average time spent in the abstinent state, or equivalently the mean time from the completion of treatment to relapse to drug use, equals $1/\iota$),

$\mu =$ the annual per treatment slot “turnover rate” (so the average time spent in drug treatment equals $1/\mu$ per treatment episode),

$p =$ the fraction of treatment episodes that are successfully completed (and hence return the treated drug user to the abstinent state; the fraction $1 - p$ of treatment episodes that fail return the drug user to the “using drugs and not waiting for treatment” state),

$\alpha =$ the annual per capita treatment request rate from active drug users not waiting for treatment (so the average time from initiation of drug use following abstinence to a request for or remand to treatment equals $1/\alpha$), and

$\delta =$ the annual dropout rate per drug user waiting in the treatment queue (so the average time a drug user is willing (or able) to wait for treatment, the *tolerance for delay* in the text, equals $1/\delta$; note that drug users waiting for treatment who commit crimes or parole violations can be removed from the queue).

These equations solve to yield

$$a(t) = \frac{s\mu p}{\iota}\bigl(1 - \mathrm{e}^{-\iota t}\bigr) + a(0)\mathrm{e}^{-\iota t} \tag{A.3}$$

and

$$\begin{aligned} q(t) = {} & \left(\frac{[n - s(1 + \mu p/\iota)]\alpha - s\mu}{\alpha + \delta}\right)\bigl(1 - \mathrm{e}^{-(\alpha+\delta)t}\bigr) \\ & + q(0)\mathrm{e}^{-(\alpha+\delta)t} - \frac{\alpha}{\alpha + \delta - \iota}\bigl(a(0) - s\mu p/\iota\bigr) \\ & \times \bigl(\mathrm{e}^{-\iota t} - \mathrm{e}^{-(\alpha+\delta)t}\bigr), \end{aligned} \tag{A.4}$$

where $a(0)$ and $q(0)$ are the initial numbers of abstinent drug users and users waiting for treatment respectively. In particular, equation (A.4) reports the number of individuals waiting in queue to enter drug treatment programs over time.

A.1. Waiting times and service levels

A.1.1. Waiting times and the forward queue

Suppose a drug user has just requested (or been remanded to) drug treatment, and encounters a waiting list of size q. There are two important questions to answer. First, how long will it take to deplete the waiting list? Assuming that users enter treatment in first-come first-served fashion, the answer to this question corresponds to the waiting time required to enter treatment. Second, what is the probability that this user will actually enter a treatment program? This *service level* corresponds to the probability that the user does not drop out of the treatment queue.

To solve for the waiting time, let $f_q(t)$ correspond to the remaining number of waiting users in front of a person t time units following his or her treatment request, when the initial queue encountered was of size q. We refer to $f_q(t)$ as the *forward queue*. Note that by definition, $f_q(0) = q$. The waiting time required to enter treatment given an initial queue of length q, denoted by $w(q)$, is then defined by the equation

$$f_q\bigl(w(q)\bigr) = 0 \tag{A.5}$$

for following $w(q)$ time units, the original forward queue will have depleted. To solve this equation, we first must produce an expression for $f_q(t)$. Noting that since the forward queue depletes due to admission to treatment (with rate $s\mu$) or to dropout (with rate $\delta f_q(t)$), we obtain the equation

$$\frac{\mathrm{d}f_q(t)}{\mathrm{d}t} = -s\mu - f_q(t)\delta \tag{A.6}$$

which, combined with the initial condition $f_q(0) = q$ noted above, solves to yield

$$f_q(t) = \left(\frac{s\mu}{\delta} + q\right)\mathrm{e}^{-\delta t} - \frac{s\mu}{\delta}. \tag{A.7}$$

Equation (A.7) can now be used to solve equation (A.5), which yields

$$w(q) = \frac{1}{\delta}\log\left(\frac{s\mu + \delta q}{s\mu}\right). \tag{A.8}$$

A.1.2. Service levels

What is the likelihood that a person who requests drug treatment when q persons are already in the treatment queue will survive (that is, *not* drop out) and actually enter treatment? Note that our assumption of a constant dropout rate δ is equivalent to assuming that the tolerance for delay is exponentially distributed with a mean duration of $1/\delta$. From this exponential distribution, the probability that a user is willing (or able) to survive at least t time units in the queue, $S(t)$, is given by

$$S(t) = \mathrm{e}^{-\delta t}, \tag{A.9}$$

as is well known [19]. We have already shown that the time required to enter treatment given an initial waiting list of length q is given by $w(q)$ from equation (A.8). Consequently, the probability that a user actually receives treatment upon encountering a waiting list of length q, $p(q)$, is given by the probability of surviving a wait of duration $w(q)$, which is equal to $S(w(q))$. We have thus demonstrated that the service level is given by

$$p(q) = S\bigl(w(q)\bigr) = \mathrm{e}^{-\delta w(q)} = \frac{s\mu}{s\mu + \delta q}. \tag{A.10}$$

Figure 2 in the text plots service levels versus waiting times for different values of δ.

Note that equations (A.8) and (A.10) are correct in both the short and the long run. As conditional statements, they only require the length of the treatment queue, q, encountered at the time a new treatment request arrives. Steady state waiting times and service levels result via inserting the steady state queue length ($\overline{q}(s)$ of equation (A.11) below) into equations (A.8) and (A.10).

A.2. Determining the capacity required for treatment on demand

A.2.1. The long run (steady state)

Treatment on demand corresponds to instantaneous entry for all users who request admission (or are remanded) to a treatment program. Equivalently, treatment on demand implies that those requesting treatment will not have to wait in a treatment queue, a circumstance that is only possible if the treatment queue is eliminated. The policy question then becomes: what is the smallest number of treatment slots required to eliminate the waiting list?

In the long run (or *steady state*), this question is easily answered. From equation (A.4) above, it is clear that over time, the length of the treatment queue as a function of the number of treatment slots approaches the steady state value

$$\overline{q}(s) = \frac{[n - s(1 + \mu p/\iota)]\alpha - s\mu}{\alpha + \delta}. \tag{A.11}$$

This equation was used to produce figure 3 in the text. The critical number of treatment slots s^* that are required to eliminate the queue over the long run sets $\overline{q}(s^*) = 0$. Via equation (A.11) we obtain the result

$$s^* = n \times \frac{\alpha\iota}{\alpha\iota + \alpha\mu p + \iota\mu}. \tag{A.12}$$

Note that s^* is increasing in three parameters: n, α and ι. That the number of slots needed to eliminate the queue increases with the size of the drug using population is not surprising. The critical slot capacity also increases with the demand for treatment (α), and with the initiation rate of drug use from the abstinent state (ι). This last point is perhaps better understood as follows: if the average drug free time spent by those who have successfully completed treatment decreases, then the total number of active drug users will increase, which in turn will require more treatment slots to eliminate the queue. The critical value s^* *decreases* if the turnover rate per slot (μ) or the probability of successfully completing a treatment program (p) increases. It is sensible that the more successful treatment episodes become, the lower the demand for treatment will be as there will be fewer active drug users to demand treatment. Greater turnover per treatment slot simply serves to pull waiting users from the treatment queue at a quicker pace. Note that the number of slots required to eliminate the queue is independent of the dropout rate δ.

It is interesting to compare the critical slot capacity s^* to a "naïve" policy that proposes to increase the existing number of slots from s to $s + \overline{q}(s)$, with the idea that since $\overline{q}(s)$ users are waiting (in the long run) when s slots are available, increasing the number of slots in accord with the size of the queue will result in treatment on demand. Via equations (A.11) and (A.12), it is clear that the naïve guess will assign too many treatment slots whenever the condition

$$\frac{1}{\delta} > \frac{1}{\mu} \times \frac{\iota}{\iota + \alpha p} \tag{A.13}$$

is satisfied. Note that if the average time users are willing to remain in the treatment queue ($1/\delta$) exceeds the average duration of a drug treatment episode ($1/\mu$), this condition will obtain. If inequality (A.13) is false, however, then the naïve policy will provide too few treatment slots to eliminate the queue.

The capacity required to achieve weaker definitions of treatment on demand can also be determined. For example, suppose that the policy requires all those who receive treatment to do so within some time w^* after requesting treatment. The smallest number of slots that meets this condition would result in waiting times equal to w^*, and via equation (A.10) a service level of

$$p^* = \mathrm{e}^{-\delta w^*}. \tag{A.14}$$

Inverting equation (A.10) yields an implied queue length q^* of

$$q^* = \left(\frac{1}{p^* - 1}\right)\frac{s\mu}{\delta} \tag{A.15}$$

and equating this to the steady state queue length of equation (A.11) and solving for s results in the required number of slots to guarantee the initial waiting time w^* specified in the policy, $s^*(w^*)$ as

$$s^*\big(w^*\big) = s^* \Big/ \left(\frac{(\mathrm{e}^{\delta w^*} - 1)\mu/\delta}{n\alpha/(\alpha + \delta)} s^* + 1\right), \tag{A.16}$$

where s^* is the capacity required to eliminate the treatment queue. Note that $s^*(w^*)$ is smaller than s^* for any value of w^*, showing that weakening the operational definition of treatment on demand leads to a reduction in the required number of slots. And of course, if w^* is set equal to zero, the required number of slots reverts back to s^* as it should. Equation (A.16) was used to produce figure 4 in the text.

A.2.2. The short run (transient)

Though providing s^* treatment slots would enable instantaneous entry to treatment in the long run, this number of slots will not eliminate treatment queues in the short run. Indeed, depending upon the initial distribution of drug users over the four states of behavior considered in our model, the time required to eliminate the treatment queue using s^* slots could be considerable. It is therefore of interest to determine the number of treatment slots required to provide treatment on demand at all times, and not simply over the long run.

Let $s^*(t)$ denote the capacity required to eliminate the treatment queue at all times. Assuming that s_0 slots are available just before treatment on demand is implemented, the naïve policy of adding enough slots to immediately eliminate the existing queue of size $q(0)$, that is, of setting

$$s^*(0) = s_0 + q(0), \tag{A.17}$$

is no longer naïve, but in fact is absolutely required. Capacities smaller than $s^*(0)$ would allow an initial queue to remain following the launch of the policy, while capacities larger than $s^*(0)$ would waste resources. We also know that $s^*(t) \to s^*$ of equation (A.12) as t becomes large. But what happens at intermediate times?

To answer this question requires reformulating our model to *force* the treatment queue to equal zero at all times. This can only happen if all treatment requests (or remands) *immediately* receive treatment, which in turn requires modeling the number of slots $s^*(t)$ that enables instantaneous service over time. Keeping all other model assumptions (and notation) the same, the "transient treatment on demand" model can be expressed as

$$\frac{\mathrm{d}a(t)}{\mathrm{d}t} = s^*(t)\mu p - a(t)\iota \tag{A.18}$$

and

$$\frac{\mathrm{d}s^*(t)}{\mathrm{d}t} = \big(n - a(t) - s^*(t)\big)\alpha - s^*(t)\mu. \tag{A.19}$$

These equations solve to yield

$$s^*(t) = s^* + c_1\mathrm{e}^{-\lambda_1 t} + c_2\mathrm{e}^{-\lambda_2 t}, \tag{A.20}$$

where s^* is given by equation (A.12),

$$\lambda_1 = \frac{(\alpha + \iota + \mu) - \sqrt{(\alpha + \iota + \mu)^2 - 4(\alpha\iota + \alpha\mu p + \iota\mu)}}{2}, \tag{A.21}$$

$$\lambda_2 = \frac{(\alpha + \iota + \mu) + \sqrt{(\alpha + \iota + \mu)^2 - 4(\alpha\iota + \alpha\mu p + \iota\mu)}}{2} \tag{A.22}$$

and c_1 and c_2 solve the equations

$$c_1 + c_2 = s^*(0) - s^*, \tag{A.23}$$

$$\lambda_1 c_1 + \lambda_2 c_2 = s^*(0)\mu - \big(n - a(0) - s^*(0)\big)\alpha. \tag{A.24}$$

Equations (A.20)–(A.24) were used to produce figure 5 in the text.

A.3. Benefits and costs of drug treatment slots

What are the likely costs and benefits of changing slot capacity? The costs correspond to the expenses of providing additional treatment resources, and as such are proportional to the number of slots provided. As for the benefits, changing treatment capacity will alter the balance between the number in the population who are not currently using drugs ($a(t) + s$), and the number who are ($n - a(t) - s$), resulting in gains (if treatment capacity is increased) or losses (if treatment slots are removed) in the aggregate annual number of drug free years in the population. To the extent that important societal measures such as crime, employment, infectious disease transmission, and child abuse are different for those who use drugs versus those who do not, gains (or losses) in such measures can be expected to accompany changes in aggregate drug free years.

A.3.1. Steady state results

Starting with a steady state calculation, let $\overline{u}(s)$ denote the steady state number of active users when there are s drug treatment slots, and $\overline{\imath}(s)$ denote the steady state number of inactive drug users. Assuming that those who are abstinent or in drug treatment are not actively using drugs, while all else in the population are, we have

$$\overline{\imath}(s) = \overline{a}(s) + s \tag{A.25}$$

and

$$\overline{u}(s) = n - \overline{\imath}(s), \tag{A.26}$$

where $\overline{a}(s)$ corresponds to the steady state number of abstinent users, and equals (by letting $t \to \infty$ in equation (A.3))

$$\overline{a}(s) = \frac{s\mu p}{\iota}. \tag{A.27}$$

Substituting equation (A.27) into equation (A.25), we see that

$$\overline{\imath}(s) = s\left(1 + \frac{\mu p}{\iota}\right). \tag{A.28}$$

Since $\overline{\imath}(s)$ also equals the steady state annual number of drug free years achieved, we see that the marginal gain in drug free years per incremental treatment slot equals

$$\frac{\mathrm{d}\overline{\imath}(s)}{\mathrm{d}s} = 1 + \frac{\mu p}{\iota}. \tag{A.29}$$

This can be understood as the sum of direct and indirect effects. For each treatment slot added, there is a direct gain of one drug free year due to the occupancy of the new treatment slot. However, there is also an indirect gain of $\mu p/\iota$ years, which can be thought of as the incremental abstinence brought about due to successful treatment completions from the incremental slot. This is the basis for our conclusion in section 4.4 that each incremental treatment slot brings about, in the steady state, an additional 3.78 drug free years (for in addition to the one year direct effect, the data suggest that $\mu p/\iota = 2.78$ years). Equation (A.29) remains valid providing the number of slots remains below s^*, the steady state capacity required to eliminate the queue (equation (A.12)).

A.3.2. Transient results

While the direct benefits of expanding treatment capacity will be visible in the short run, it would take time for the indirect effects to manifest. Over a time horizon running from 0 to τ, the annual average number of drug free years in the population as a function of system capacity, $\overline{\imath}(s,\tau)$, equals

$$\begin{aligned}\overline{\imath}(s,\tau) &= \frac{1}{\tau}\int_0^\tau \bigl(a(t) + s\bigr)\,\mathrm{d}t \\ &= s\left(1 + \frac{\mu p}{\iota}\right) + \left(a(0) - \frac{s\mu p}{\iota}\right)\frac{1 - \mathrm{e}^{-\iota\tau}}{\iota\tau}.\end{aligned} \tag{A.30}$$

The incremental gain in annual drug free years per treatment slot added is thus equal to

$$\frac{\partial\overline{\imath}(s,\tau)}{\partial s} = 1 + \frac{\mu p}{\iota} - \frac{\mu p}{\iota}\frac{1 - \mathrm{e}^{-\iota\tau}}{\iota\tau}. \tag{A.31}$$

This result shows that over very short time horizons, the incremental gain in annual drug free years per treatment slot added equals 1, the direct benefit of placing another user into treatment. As the time horizon τ increases, however, the indirect effect of additional drug free years on account of successful treatment episodes grows until the full indirect benefit of $\mu p/\iota$ drug free years per treatment slot is reached. A graph of equation (A.31) appears as figure 6 in the text.

A.4. Parameter estimates used in the text

The parameters employed in the case study reported in section 4 were derived by combining empirical observations from San Francisco and other national studies with the assumption that the treatment system in San Francisco is in steady state. This latter assumption imposes consistency conditions for determining otherwise unknown parameter values.

The parameter values used are shown in table 1 of the text. Note that n, $q(0)$, s, μ, p, and ι were assigned empirically observed values. The initial number of abstinent users was assigned via the steady state assumption following equation (A.27) (that is, $a(0)$ was set equal to $\overline{a}$). The two remaining parameters δ and α can be related using the steady state assumption: since $\mathrm{d}q(t)/\mathrm{d}t = 0$ in the steady state, equating the left hand side of equation (A.2) to zero and solving for α in terms of δ yields the steady state relation

$$\begin{aligned}\alpha &= \frac{s\mu}{n - s - \overline{a} - \overline{q}} + \frac{\overline{q}\delta}{n - s - \overline{a} - \overline{q}} \\ &= 1.127 + 0.071\delta\end{aligned} \tag{A.32}$$

after inserting the values of the other parameters shown in table 1 (where $\overline{q} = q(0)$ following the assumption that the observed system was already in steady state).

We are left with one free parameter, the dropout rate δ. In our examples, we have considered dropout rates that imply a tolerance of delay of 1 year, 1 month, 1 week and 1 day (corresponding to δ's of 1/yr, 12/yr, 52/yr and 365/yr). Each of these different δ values produce corresponding α's via equation (A.32), resulting in the four scenarios explored in section 4. For those examples in section 4 where the initial queue was set to 1,000 or 500 as opposed to 1,400, the appropriate value was substituted for $\overline{q}$ in equation (A.32) to produce the appropriate demand for treatment to use in the analysis.

References

[1] D.R. Gerstein and H.J. Harwood, eds., *Treating Drug Problems* (National Academy Press, Washington, DC, 1990).

[2] Substance Abuse and Mental Health Services Administration, *National Treatment Improvement Evaluation Study* (Department of Health and Human Services, Center for Abuse Treatment, Washington, DC, 1996).

[3] C.P. Rydell, J.P. Caulkins and S.S. Everingham, Enforcement or treatment? Modeling the relative efficacy of alternatives for controlling cocaine, Operations Research 44 (1996) 687–695.

[4] D.R. Gerstein, R.A. Johnson, H.J. Harwood, D. Fountain, N. Suter and K. Malley, *Evaluating Recovery Services: The California Drug and Alcohol Treatment Assessment, General Report* (National Opinion Research Center, Chicago, 1994).

[5] National Association of State Alcohol and Drug Abuse Directors, *Estimated Number of Individuals Needing Treatment* (National Association of State Alcohol and Drug Abuse Directors, Washington, DC, 1996).

[6] United States General Accounting Office, *Drug Treatment: Some Clinics Not Meeting Goal of Prompt Treatment for Intravenous Drug Users* (Publication No. GAO/HRD-90-80BR, General Accounting Office, Washington, DC, 1990).

[7] S. Shane, Cost of addiction carries "hidden tax": city doubles spending on treatment to cut price paid by public, Baltimore Sun (April 30, 1998) 1A.

[8] D. Weikel, A grueling waiting game for addicts seeking help: drug treatment is highly cost-effective, data show, but programs for the indigent fail to meet enormous demand, Los Angeles Times (April 24, 1997) 1.

[9] W. King, Prompt treatment key to recovery – but detox programs now in country aren't large enough to meet demand, Seattle Times (January 21, 1998) A12.

[10] C. Bowman, $20 million plan to help S.F. addicts – health officials back "treatment on demand", San Francisco Chronicle (November 8, 1996) A21.

[11] D.C. Des Jarlais, Personal communication to Edward H. Kaplan via e-mail (May 12, 1998).

[12] R.C. Bailiey, Y.-I. Hser, S.-C. Hsieh and M.D. Anglin, Influences affecting maintenance and cessation of narcotics addiction, The Journal of Drug Issues 24 (1994) 249–272.

[13] H.-I. Hser, K. Yamaguchi, H. Chen and M.D. Anglin, Effects of interventions on relapse to narcotics addiction: an event-history analysis, Evaluation Review 19 (1995) 123–140.

[14] M.G. Dekimpe, L.M. Van De Gucht, D.M. Hanssens and K.I. Powers, Long-run abstinence after narcotics abuse: what are the odds?, Management Science 44 (1998) 1478–1492.

[15] San Francisco Department of Public Health, *Supporting the Department of Public Health's Substance Abuse Treatment on Demand: First Steps Plan to Guide the Augmentation of Alcohol and Drug Services in Fiscal Year 1997–98, May 20, 1997* (retrieved via http://www.dph.sf.ca.us/hcres/html/hc_resolutions_52.htm on January 26, 1999).

[16] M. Schlesinger and R. Dorwart, Falling between the cracks: failing national strategies for the treatment of substance abuse, Daedelus 121 (1992) 195–237.

[17] C. Marinucci, Brown wrong about drug treatment, advocates say, San Francisco Chronicle (October 17, 1997) A21.

[18] H.J. Harwood, R.L. Hubbard, J.J. Collins and J.V. Rachal, The costs of crime and the benefits of drug abuse treatment: a cost-benefit analysis using TOPS data, NIDA Research Monograph 86 (1988) 209–235.

[19] J.D. Kalbfleisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data* (Wiley, New York, 1980).

[20] Substance Abuse and Mental Health Services Administration, *National Household Survey on Drug Abuse: Population Estimates 1994* (Substance Abuse and Mental Health Services Administration, Washington, DC, 1994).

[21] J.P. Walters, *Prepared Testimony of John P. Walters, Former Acting Director and Deputy Director for Supply Reduction, Office of National Drug Control Policy, Before the House Committee on International Relations*, April 29 (Federal News Service, Washington, DC, 1998).