Probability Models of Needle Exchange

Edward H. Kaplan

Yale University, New Haven, Connecticut

(Received December 1993; revision received July 1994; accepted August 1994)

Needle exchange is an intervention to slow the spread of HIV infection among drug users (IDUs) via needle sharing (Stimson 1989, Des Jarlais and Friedman 1992). Whether needle exchange achieves this end is hard to determine via direct observation. To conclusively measure new HIV infections requires repeated HIV tests of IDUs participating in outreach programs; such studies are rare (for an exception see Ljungberg et al. 1991). There have been no controlled trials pitting HIV incidence among IDUs participating in a needle exchange against incidence in a provably equivalent group of IDUs not exchanging needles.

Given these difficulties, researchers have turned to behavioral studies of needle exchange participants. Such studies rely on self-reported changes in risky behaviors such as needle sharing as evidence of the impact of needle exchange (Donoghue et al. 1989, Harrigan et al. 1989, Joseph and Des Jarlais 1989, Watters and Cheng 1991, Hagan et al. 1991, GAO 1993). Social desirability effects might skew self-reported behavioral data. In a program environment where participants are instructed not to share needles, what should one expect in response to questions like "How often do you share needles?" This does not prove that self-reports are inaccurate, but it is hard to know the degree to which self-reported changes in behavior reflect the truth.

This paper describes a different set of ideas that have been developed for needle exchange evaluation. The common thread is the utilization of probability models of needle exchange operations in conjunction with objectively observable data, such as needle distribution rates, dates of client visits, needle circulation times, and the fraction of returned needles testing HIV positive. These models thus differ from earlier work which did require the use of client self-reports (Kaplan and O'Keefe 1993). Two different models will be described. The first estimates the incidence of HIV infection among participating IDUs using only data describing the dates of client visits and the measured level of infection in needles following Kaplan and Heimer (1994). The second approach estimates the absolute reduction in new HIV infections due to needle exchange by extending earlier models of relative impact (Kaplan 1994). Both models have been implemented using data from the legal needle exchange program operated by the AIDS Division of the New Haven, Connecticut Department of Health. The models suggest that needle exchange has reduced HIV transmission among program participants from the start of the program in November 1990 through June 1992.

In July 1992, the purchase of syringes without prescription at Connecticut pharmacies became legal (Valleroy et al. 1993, Weinstein and Hadler 1993). Concomitant with the decriminalization of syringe purchase in Connecticut was a drop in the number of IDUs participating in the New Haven program. As decriminalization changed the environment in which the New Haven program was operating, most of the data discussed in this paper will be limited to the first 20 months of the program.

The remainder of this paper is organized as follows. Section 1 describes the first model (based on change point theory) and its results, while Section 2 describes the second model (based on a circulation theory of needle exchange) and its results. The role this research has played in public health policy will be discussed briefly in Section 3.


Operations Research
Vol. 43, No. 4, July-August 1995

558

0030-645X/95/4304-0558 $01.25 © 1995 INFORMS
1. MODEL 1: MAXIMUM LIKELIHOOD CHANGE POINT MODEL

1.1. An Idealized Research Environment

Imagine a research environment where one can repeatedly test a cohort of $m$ IDUs for HIV infection over time. If the prevalence of HIV infection in the IDUs is denoted by $\phi$ and the incidence rate of new infections is given by $\rho$ infections per uninfected IDU per unit time, then the probability that an IDU would become infected over a duration of length $\delta$ is given by

$$P(I_{\text{Infection}}) = (1 - \phi)(1 - e^{-\rho\delta}) \approx (1 - \phi) \rho \delta \equiv \delta \phi$$

(1)

for sufficiently small values of the product $\rho \delta$. Note that this is a short run perspective, in that it is assumed that both the prevalence $\phi$ and the incidence $\rho$ remain constant over the duration of the study.

Let $y_i = 1$ if the $i$th IDU in the sample becomes infected during $\delta$, and $y_i = 0$ otherwise. From data, one would discover that approximately $m \phi$ IDUs in the cohort were infected at the start of the study (for the prevalence equals $\phi$ by assumption). Second, one would reason that $\sum y_i = m(1 - \phi) \rho \delta$ based on (1). The raw incidence rate $\rho$ could thus be estimated by dividing $\sum y_i$ by $m(1 - \phi) \rho \delta$ (actually, one would divide $\sum y_i$ by the product of $\delta$ and the observed number of initially uninfected IDUs, but this latter number is roughly $m(1 - \phi)$). At the end of duration $\delta$, one would have two ways of expressing the incidence rate: the number of new infections per uninfected drug injector per unit time ($\rho$), or the unconditional number of new infections per IDU per unit time ($\delta \phi$). It is easy to show that the expression $\sum y_i/m$ is also the maximum likelihood estimator of $\delta$ (again assuming that $\rho \delta$ is sufficiently small).

1.2. A Needle Exchange Environment Where Only Needles Are Tested

With the above computations in mind, consider the more difficult research environment of a needle exchange. Suppose that the only data available report the dates of client visits and the HIV status of returned needles. Specifically, suppose that for a given IDU the following data are available: the earliest and latest dates of needle exchange (and, hence, the duration $\delta$ that the IDU was exposed to the program), and a series of indicator variables $x_1, x_2, \ldots, x_n$ corresponding to $n$ returned needles ordered by date of return. The indicator $x_i = 1$ if the $i$th needle in sequence tests positive, and zero otherwise. What can one infer from such data?

Note that if needles were never shared and all laboratory testing of needles was error free, the above data would yield information equivalent to the idealized incidence trial described above. However, in the presence of such distractions, it is still possible to estimate statistically whether a given IDU has become infected. This is because of the following observation: whatever the laboratory error rates and sharing patterns are, the likelihood that a returned needle will test positive for HIV after an IDU has become infected should be higher than the likelihood that a needle returned by the same IDU would test positive before the IDU became infected. One should, therefore, be able to detect a new infection by observing an upward shift in the fraction of returned needles that test positive.

1.3. A Maximum Likelihood Change Point Model

This intuition can be formalized. Let $\pi(-)$ and $\pi(\cdot)$ denote the probability that needles test positive conditional on their return by an IDU who is infected (uninfected). As this procedure is being applied on an IDU by IDU basis, there is no need for the values of $\pi(\cdot)$ and $\pi(-)$ to be common across IDUs. Indeed, these probabilities allow for differences in needle sharing and needle cleaning patterns across IDUs, in addition to pure laboratory testing errors. However, for a given IDU, these probabilities are assumed to remain constant over the period of study. This latter assumption is equivalent to assuming stability in needle sharing and cleaning practices (as well as laboratory error), and seems reasonable for relatively short time durations.

Suppose that the IDU in question became infected between the $s$th and $s + 1$st needles tested, and that knowledge of $s$ yields no information regarding the values of $\pi(-)$ and $\pi(\cdot)$. The probability of observing the data $x_1, x_2, \ldots, x_s, x_{s+1}, \ldots, x_n$ is then given by the likelihood function

$$L = \prod_{i=s+1}^{n} \pi(-)^{(1 - \pi(-))^{x_i}} \cdot \prod_{i=s+1}^{n} \pi(\cdot)^{\pi(\cdot)^{x_i}} \cdot (1 - \pi(\cdot))^{1-x_i}$$

(2)

The model formulated above is a change point model (Page 1955, Cox 1970), and the unknown parameters $\pi(-)$, $\pi(\cdot)$, and $s$ can be estimated from the data record $x_1, x_2, \ldots, x_n$ via the method of maximum likelihood. Conditioned on $s$, it is clear that the maximum likelihood estimators for $\pi(-)$ and $\pi(\cdot)$ are given by $\hat{\pi}(-) = x_s/s$ and $\hat{\pi}(\cdot) = x_{s+1}/(n - s)$ unless the former is larger than the latter, in which case the maximum likelihood estimators for both probabilities are given by $\hat{\pi}(\cdot) = x_{s+1}/(n - 1)$ for logic insists that $\pi(\cdot) \neq \pi(-)$. Substituting the above formulas into (2) enables one to conduct a search over $s \in \{1, 2, 3, \ldots, n - 1\}$ to find that value of $s$ that maximizes the value of $L$ in (2) (actually, one only needs to consider those values of $s$ such that $x_s = 0$), in the search process, as is easily shown; call the maximized value of the likelihood function $L_M$.
infected or because the IDU was uninfected and remained so over the study period). Under the null hypothesis, the maximum likelihood estimator for \( \pi \) is given by \( \sum i=1, n \) \( x_i/n \). Substituting this value for both \( \pi(-) \) and \( \pi(+) \) in (2) yields the maximized likelihood under the null hypothesis; call this \( L_0 \).

A threshold test can be established as follows: Reject the null hypothesis, and conclude that the IDU became infected, if the ratio \( \hat{L} / L_0 \) exceeds some cutoff denoted by \( c \). Kaplan and Heimer (1994) demonstrate numerically that selecting \( c = 5.991 \) yields a type-I error probability of \( \alpha = 0.05 \) and a type-II error of \( \beta \leq 0.50 \) for reasonable combinations of the parameters \( \pi(-), \pi(+) \), and \( s \).

1.4. Estimating HIV Incidence From Change Point Test Results

Recall from (1) that the probability that an IDU becomes infected over some duration \( \delta \) approximately equals \( \theta \delta \), where \( \delta \) is the number of new infections per IDU per unit time. Instead of testing individual IDUs, imagine observing the results of change point tests conducted on the sequences of needles returned by program participants. What is the probability that one would conclude an IDU became infected on the basis of a change point test?

There are two possibilities. First, the IDU in question really became infected (with probability \( \theta \delta \)), and the change point test had the power to detect this new infection (with probability \( 1 - \beta \)). Alternatively, the IDU did not become infected (with probability \( 1 - \theta \delta \)), but the change point test committed a type-I statistical error, and falsely concluded that the IDU became infected (with probability \( \alpha \)). Therefore, the probability \( r(\theta) \) that one would conclude an IDU became infected on the basis of a change point test is given by

\[
r(\theta) = \theta \delta (1 - \beta) + (1 - \theta \delta) \alpha = \alpha + (1 - \alpha - \beta) \theta \delta.
\]

(3)

By analogy with subsection 1.1, let \( y_i = 1 \) if the change point test performed on the sequence of needles returned by the \( i \)-th IDU suggests that infection has occurred (i.e., \( 2 \log(\hat{L} / L_0) > c \) for the \( i \)-th sequence of needles), and \( y_i = 0 \) otherwise, \( i = 1, 2, \ldots, m \). A simple test of the hypothesis that no new infections have occurred among the IDUs returning the needles can be constructed by setting \( \theta = 0 \) (as required by the hypothesis of no new infections). In this case, \( r(\theta) = \alpha \) and the total number of rejected change point tests \( \sum i=1, m \) \( y_i \) follows the binomial probability distribution with parameters \( m \) and \( \alpha \). It is easy to compute the probability that at least as many rejected change point tests as actually observed would occur solely due to chance.

It is also possible to estimate \( \theta \) via maximum likelihood. Given \( \delta \), the duration of exposure of the \( i \)-th IDU to the needle exchange calculated from the dates of first and last program visits (or, perhaps, more appropriately, from the dates of first and last visits from which returned needles were tested for HIV), the probability of observing a rejected change point test follows (3) after substituting \( \delta \) for \( \delta \). As a consequence, the probability of observing the particular set of change point results for all \( m \) IDUs is given by

\[
L(\theta) = \prod i=1, m \left( r(\theta) \delta \right)^{y_i} \left( 1 - r(\theta) \right)^{1-y_i}.
\]

(4)

Maximizing \( L(\theta) \) over \( \theta \) subject to the constraint that \( \theta \geq 0 \) can be achieved directly. The maximum likelihood estimator \( \hat{\theta} \) will equal zero for any change point significance level \( \alpha \geq \sum i=1, m \) \( y_i / \sum i=1, m \) \( \delta i = \delta \); otherwise, \( \hat{\theta} \) is the unique root of the equation

\[
\sum i=1, m \ y_i \delta / r(\hat{\theta}) = \sum i=1, m \ (1 - y_i) \delta / (1 - r(\hat{\theta})).
\]

(5)

Note that if \( \delta_i = \delta \) for \( i = 1, 2, \ldots, m \), then (5) yields a simple formula for \( \hat{\theta} \), given by

\[
\hat{\theta} = \frac{\sum i=1, m \ y_i - m \alpha}{(1 - \alpha - \beta) m \delta}.
\]

(6)

In the idealized world of perfect statistical testing, both \( \alpha \) and \( \beta \) would equal zero, yielding \( \hat{\theta} = \sum i=1, m \ y_i / m \delta \), the same result cited in subsection 1.1.

1.5. Application to the New Haven Needle Exchange: The Syringe Tracking and Testing System and Statistical Results

The data structures alluded to above have been realized as part of the evaluation of New Haven's needle exchange program. The unique syringe tracking and testing system (or STT) implemented in New Haven works as follows (Kaplan and O'Keeffe): All program participants receive a code name of their own choosing, and all needles distributed by the program are given a unique tracking code. When exchange transactions occur, outreach workers record the date, location, and code name of the person receiving the needles. When needles are returned, the date, location, and code name of the person returning the needles are also recorded. Consequently, a complete record of syringe transactions and client participation is compiled.

In addition, a sample of returned needles is tested for the presence of HIV (Heimer et al. 1992). Since a simple random sample of returned needles would be weighted heavily to reflect those who exchange most often, a two-stage sampling procedure has been implemented to achieve greater client coverage. Needles are exchanged in batches of at most five. In the first sampling stage, all syringes are retrieved from returned batches of at most two, while two syringes are taken from batches of size three through five. The second stage systematically samples every tenth needle selected in the first stage. From
November 1990 through June 1992, this procedure selected 2,813 of the 49,405 needles distributed for testing. The change point models were implemented on a subset of the needle testing data reported above—needles returned by the end of May 1992 and only those from clients who had returned at least five tested needles, because the change point test cannot provide meaningful results for fewer than five needles. The STT provided a total of 1,920 needles representing 132 different clients with at least five tested needles (averaging 14.5 tested needles per client). Approximately 42% of these needles tested HIV positive via polymerase chain reaction (PCR). Exposure times were taken as the time between the first and last dates tested needles were returned inclusive. The total derived exposure time was 34,903 person days in the needle exchange, averaging 264 days per client.

Of the 132 change point tests conducted, six rejected the null hypothesis of no infection using the cutoff $c = 5.991$. If one were to simply divide this number by the total exposure time, one would estimate an unconditional HIV incidence rate of 6.3 new infections per 100 needle exchange clients per year. This ignores the chance errors one can expect from applying the change point test. The type-I error probability of $\alpha = 0.05$ associated with the change point test easily accounts for the six rejections observed. Under the null hypothesis that no new infections occurred, that is, that $\theta = 0$, the probability of observing six or more rejected change point tests is given by the binomial tail probability

$$P_r(\text{Number of Rejections} \geq 6) = \sum_{i=6}^{132} \binom{132}{i} 0.05^i (1 - 0.05)^{132-i} = 0.6512.$$  

With such a large tail probability, it is not possible to reject the hypothesis that in the New Haven needle exchange, no new infections have occurred among those IDUs studied ($\theta = 0$).

The maximum likelihood approach supports this point as well. The value of $\sum \delta_i \gamma_i \delta_i = 0.0598$. For any $\gamma_i \geq 0.0398$, $\delta = 0$, and as $\alpha$ is in the neighborhood of 0.05, it is indeed the case that $\theta = 0$.

These results were not preordained. It could have been the case that the change point test rejected the null hypothesis for a large percentage of the clients studied. At the conventional significance level of 5%, one would not be able to reject the null hypothesis of no new infections unless 12 or more change point tests suggested that infection had occurred. That only half that number was actually observed suggests that in the New Haven needle exchange program, the incidence rate of new infections was not high. Furthermore, since more than 42% of the needles tested HIV positive, it is clear that this population had a positive HIV incidence rate in the past.

The change point model provides an approach to estimating HIV incidence among needle exchange participants but does not explain how needle exchange operations might be expected to affect HIV transmission rates. This requires deeper understanding of the physics of needle exchange as described next.

2. MODEL 2: CIRCULATION THEORY OF NEEDLE EXCHANGE

2.1. Intuition Underlying the Circulation Theory Model

Imagine the HIV status of a needle from the moment it is introduced into a population of drug injectors. Many different experiences are possible, but at any moment in time a needle may be viewed as either contaminated or not. Needles can only become infected following use by an HIV-infected IDU, though not all needles necessarily become contaminated following such use (also, effective post-injection bleeding may render some needles noninfectious in spite of having been used by infected IDUs).

Infected needles can become decontaminated for a variety of reasons, including effective preinjection bleeding (though not all bleeding is effective), dilution of infectious blood with the blood of uninfected IDUs following several shared injections, or natural inactivation of the virus. Whatever the relative contamination and decontamination rates are per needle, clearly the longer a needle remains in circulation, the higher the likelihood that the needle becomes infected.

Aggregating the stochastic infection processes for individual needles across all needles in the population leads to an average level of infection among all needles at any point in time. This average level may be interpreted as the probability that a needle selected at random is contaminated with HIV. In addition to the relative contamination and decontamination rates, the average infection level depends on the distribution of needle circulation times, with stochastically longer circulation times associated with higher average infection levels in the needles.

Needle exchange may be thought of as a mechanism that operates directly on the distribution of needle circulation times (Kaplan and Heimer 1993, Kaplan 1994a, b).

First note that by exchanging needles, the total number of needles in circulation does not change in theory: The rate with which needles are distributed is balanced by the rate with which needles are returned. To the extent that one believes that HIV transmission among IDUs is a function solely of the number of needles in circulation, then one should not expect needle exchange to change matters much. Instead, what one achieves by exchanging needles is the interruption of the needle circulation process. This stochastically reduces the length of time for which needles remain in the population. The greater the rate with which needles are exchanged, the greater the reduction in needle circulation times. This is the key operational link between needle exchange programs and HIV transmission, for reducing needle circulation times...
acts via the stochastic infection process to lower the average level of infection in the population of needles. The intuition is simple: If needles are available for shorter periods of time per needle, then the likelihood that different IDUs will use the same needle declines. In effect, needle exchange causes needles to share fewer people.

Reducing the mean level of infection in circulating needles is important, for those who share needles are in effect sampling from the population of circulating needles. If the mean level of infection in needles declines, then the chance that a person who shares needles will encounter a contaminated needle also declines. Since one can only become infected via needle sharing from contaminated needles, it follows that the incidence rate of HIV infection by needle sharing will be proportional to the average level of infection in circulating needles. The relative success of a needle exchange in reducing HIV transmission can, therefore, be measured by the relative drop in the level of infection in circulating needles.

The theory just summarized does not imply that needle exchange is an instant success. To be effective in reducing HIV incidence appreciably, the theory suggests that exchange programs must exchange needles rapidly enough to reduce circulation times by a large amount. The validity of the theory was tested using data from New Haven's syringe tracking and testing system; doing so required more formal models, as described in subsections 2.2-2.5.

2.2. The Needle Infection Process

Circulating needles are always in one of two states, HIV-contaminated or uncontaminated. Aiming for simplicity, a two-state, continuous-time Markov process is proposed as a model for the dynamics of needle infection. Uncontaminated needles are assumed to become infected with rate \( \lambda \) per uncontaminated needle per unit time, while HIV-contaminated needles become uncontaminated with rate \( \mu \) per contaminated needle per unit time. We assume a short term perspective, so the prevalence of infection \( (\phi) \), needle injection, sharing and cleaning rates, and viral inactivation rates are assumed constant over the period of study, leading to stable values of \( \lambda \) and \( \mu \). Let \( \pi(t) \) denote the probability that a needle that has been circulating for \( t \) time units is HIV-contaminated. This can be found from the first-order linear differential equation

\[
\frac{d\pi(t)}{dt} + (\lambda + \mu)\pi(t) = \lambda
\]

which solves to yield

\[
\pi(t) = \frac{1}{\lambda + \mu} + \left( \pi_0 - \frac{1}{\lambda + \mu} \right) e^{-(\lambda + \mu)t},
\]

where \( \pi_0 \) is the probability that the needle is already HIV-contaminated when introduced to the population (and should equal zero if the needle in question was provided via a needle exchange program). Equation (9) quantifies the intuition described earlier: The probability that a needle is HIV-contaminated increases from \( \pi_0 \) to the equilibrium value of \( \lambda/(\lambda + \mu) \) with increasing circulation time.

2.3. The Needle Circulation Process

Equation (9) describes the probability that a needle is HIV-contaminated following a fixed circulation time of duration \( t \). However, there are many needles circulating at any moment in time. Of interest is the mean level of infection in circulating needles (i.e., the probability that a randomly selected circulating needle is infected).

Computing this probability requires a model of the needle circulation process. Let random variable \( T \) denote the complete length of time for which a needle remains in the population, and random variable \( T_e \) denote the elapsed circulation time for a randomly sampled circulating needle. The random variable \( T_e \) is then the equilibrium backwards recurrence time associated with length biased sampling of random variable \( T \) (Cox 1962), for the likelihood of finding a complete circulation interval of a given duration is proportional to both the frequency with which such intervals occur and the duration of the interval itself. In studies conducted thus far, we have assumed an exponential circulation time probability density function given by

\[
f_T(t) = \frac{1}{\tau} e^{-t/\tau},
\]

where \( \tau \) is the mean circulation time. The exponential possesses the special property that the equilibrium distribution of elapsed circulation times equals the original distribution of circulation times. The average level of infection in circulating needles for the Markov-exponential model, denoted by \( \bar{\pi}(\tau) \), is thus given by

\[
\bar{\pi}(\tau) = \int_0^\infty \left\{ \frac{1}{\lambda} + \frac{\mu}{\lambda + \mu} \right\} e^{-\tau} \lambda^\tau + \frac{\lambda}{\lambda + \mu} e^{-\tau}\lambda^\tau + \frac{\mu}{\lambda + \mu} e^{-\tau}\mu^\tau \frac{e^{-\tau}\lambda^\tau + \mu^\tau}{\tau^\tau} d\tau.
\]

Equation (11) has the correct intuitive properties: As \( \tau \) approaches 0, \( \pi(\tau) \) approaches \( \pi_0 \), and as \( \tau \) becomes very large, \( \bar{\pi}(\tau) \) approaches the equilibrium value \( \lambda/(\lambda + \mu) \).

Having established a model that describes the average level of infection in circulating needles, it is possible to consider how needle exchange affects this mean level of infection.

2.4. Needle Exchange Reduces Circulation Times

Let \( v \) denote the rate at which needles are exchanged per IDU per unit time, and \( \pi(v) \) denote the mean circulation time that results when needles are exchanged with rate \( v \). The mean removal rate of needles from the population when a program with exchange rate \( v \) is in force
will be denoted by $\xi(v)$. Clearly $\xi(v) = 1/\tau(v)$ by definition.

In the absence of needle exchange ($v = 0$), there will still be some baseline removal rate ($\xi(0)$) and mean circulation time ($\tau(0)$) in the population of needles. However, as $v$ increases, one expects $\xi(v)$ to increase as well. In particular, let

$$\xi(v) = a + bv$$  \hspace{1cm} (12)

represent the impact of needle exchange on the removal rate of needles (the parameters $a$ and $b$ are both nonnegative). The parameter $a$ denotes the baseline removal rate, so $\xi(0) = 1/a$, while $b$ may be interpreted as a transfer coefficient converting needle distribution per client to needle removal per circulating needle. The coefficient $b$ accounts for the difference in the population size of clients and the population size of needles (the greater the ratio of clients to needles, the greater the value of $b$).

In addition, $b$ allows for the possibility of substitution effects, in that the availability of needles from an exchange program may partially substitute for needles previously obtained from other sources (the greater the degree of substitution, the lower the magnitude of $b$).

The models above present a compact theory of the physical effect of needle exchange: Implementing a needle exchange program with an exchange rate of $v$ needles per client per unit time is hypothesized to increase per needle removal rates ($\xi(v)$) in accordance with (12). This reduces the average circulation time among circulating needles for $\tau(v)$ equals $1/\xi(v)$. Reducing $\tau(v)$ reduces $\tau(v)\gamma$ in accordance with (11).

As HIV transmission via needle sharing requires injection with a contaminated needle, it is reasonable to consider the relative reduction in the mean level of HIV infection in circulating needles as a measure of the relative reduction in HIV incidence. The key evaluation performance measure to be considered is, therefore, equal to

$$\text{Relative Reduction} = 1 - \frac{\tau(v)}{\tau(0)}.$$  \hspace{1cm} (13)

2.5. Back-of-the-Envelope Estimates of Baseline HIV Incidence Rates

Computing absolute reductions in incidence requires an estimate of the baseline HIV incidence rate before needle exchange began. Such estimates are not available via the theory advanced thus far, but can be obtained from other back-of-the-envelope calculations if actual observations from field studies are not available, which will most often be the case. The idea is not to be precise, but rather to develop a range of values that can be combined with (13) to produce upper and lower bounds on the actual number of infections averted by needle exchange.

The first suggestion borrows from an established result in epidemiology: For any disease in equilibrium (i.e., steady state), the prevalence of the disease equals the product of the incidence rate and the mean duration of infectiousness (Mausner and Kramer 1985). This is analogous to Little's law for queueing systems: The mean number of customers in the system equals the product of the customer arrival rate and the average sojourn time in the system. Here, new infections correspond to arriving customers, prevalent infections correspond to customers in the system, and the duration of infectiousness corresponds to the sojourn time in the system.

One can invoke the prevalence/incidence law if one believes that the HIV epidemic has become (locally) endemic among drug injectors. Then, an estimate of the unconditional incidence rate $\theta$ is given simply by

$$\theta = b\gamma,$$  \hspace{1cm} (14)

where $\theta$ is the prevalence of HIV infection, and $\gamma$ is the mean progression rate, and can be thought of as the reciprocal of the duration of infectiousness. In applying this relationship, ‘duration of infectiousness’ means the length of time during which an infectious IDU is behaving in a manner capable of spreading disease. AIDS epidemic modelers usually set $\gamma$ equal to the reciprocal of the mean time from HIV infection through the development of AIDS symptoms (Kaplan 1989, Anderson and May 1991), under the assumption that persons with frank AIDS do not continue to share needles or engage in unprotected sexual activity.

Some caustics apply to (14). First this relation is only sensible if one believes that the epidemic has achieved an equilibrium level of infection. This may be true for drug injectors in some areas of the United States, such as the northeastern seaboard (i.e., New York City and environs). Second (not all IDUs are infected via needle sharing, and it is only such infections that the physics of needle exchange can prevent. While it is not known precisely what percentage of new infections among IDUs are sexually acquired (Schoenbaum et al. 1989, Nelson et al. 1992), some researchers have produced estimates as high as 40% (Dr. James Kahn, UC San Francisco, personal communication). Therefore, it is reasonable to discount the result of (14) to account for possible sexual transmission. Third in employing (14), one either uses a prevalence estimate among program IDUs at baseline, or prevalence information from other sources. If the prevalence value used is a community value, then there may be selection effects of unknown form if those participating in the needle exchange are not representative of all IDUs in the relevant population. Some might argue that those likely to be engaged in a needle exchange are those most likely to look after their own health, an argument for reducing the prevalence among program clients relative to a community value. Alternatively, others might note that needle exchanges seem to attract older drug injectors, an argument for raising the prevalence relative to a community value. Rather than systematically attempting to sort these biases out, the suggestion offered here is to use (14) to create a range of plausible
values for which credible arguments exist. An example of this will be illustrated in subsection 2.6.

A second suggestion for estimating baseline incidence is to apply the well-known backcalculation technique for reconstructing aggregate HIV incidence data from reported AIDS incidence data corrected for reporting delays (Brookmeyer and Gail 1988, Brookmeyer 1991). Letting \( \pi(t) \) denote the AIDS incidence rate at time \( t \) (measured in the number of cases per unit time), \( \lambda(t) \) denote the HIV incidence rate at time \( t \) (measured in the number of infections per unit time), \( I \) denote the random variable representing the time needed to progress from HIV infection through AIDS, and \( f_I(t) \) denote the probability density function of \( I \) (assumed to be known), the backcalculation model is given by

\[
\pi(t) = \int_0^t \lambda(u) f_I(t - u) \, du.
\]

The logic of backcalculation is simple: For \( \pi(t) \) AIDS cases to have been reported at time \( t \), how many HIV infections must have occurred at earlier points in time?

Backcalculation can be applied to AIDS incidence data specific to IDUs to obtain a rough idea of HIV incidence rates in that population. Backcalculation offers little information about recent HIV incidence due to the long incubation period of the virus; for a critical review of the backcalculation method, see Bacchetti, Segal and Jewell (1993). However, one can still develop an idea about the magnitude of HIV incidence from backcalculation. For example, one might conjecture an HIV incidence function with a constant incidence rate for the past several years, and use backcalculation to obtain an estimate of this incidence rate.

Given an incidence estimate among IDUs obtained via backcalculation, it must be remembered that this estimate pertains to the IDU population as a whole, and also accounts for all possible sources of infection. Thus, one first needs to ask what percentage of the incidence estimated can be assigned to needle exchange clients at baseline. If one believes that clients are representative of all IDUs in the community, then one could multiply the estimated incidence rate by the ratio of the number of program participants to the number of IDUs in the community. This assumes that one has an estimate of the number of IDUs in the community, a number which may be hard to come by. The resulting incidence figure might be adjusted up or down depending upon one's beliefs regarding the likelihood of infection among needle exchange clients relative to IDUs at large as discussed above, and should be discounted to account for HIV infections acquired via sources other than needle sharing.

Although there are many uncertainties operating here, recall that the idea is not to produce a precise estimate of HIV incidence, but rather a range of plausible values. While one may not know the exact number of IDUs in a city, for example, one can reasonably bound the number using data sources such as arrest records, drug treatment data, and mortality data (Newmeyer 1988, Aldrich, Mandel and Newmeyer 1990, Frischer et al. 1993), as well as common sense (e.g., is it reasonable to believe that more than 5% of a city's population regularly injects drugs?). The approach will be illustrated in subsection 2.6.

2.6. Application to the New Haven Needle Exchange Program

The ideas of subsections 2.1-2.3 were applied to the evaluation of the New Haven needle exchange program. The sampling procedure for obtaining needles for testing was as described in subsection 1.5, and the fraction of tested needles distributed in a given month that subsequently tested HIV positive via PCR is taken as an empirical measure of \( \pi \), the mean level of infection in circulating needles among program clients. Kaplan (1994b) shows that defining \( \pi \) in this manner remains reasonable even in the presence of laboratory testing errors. As the exchange program will accept any needles in exchange for new ones, there is no reason to believe that those needles returned represent a biased selection of all needles in circulation.

Needle distribution rates per client \( i \) were defined monthly as the number of needles distributed by the program in a given month divided by the number of clients who visited the program at least once during the month in question. Mean needle circulation times were also indexed by month of distribution. As needles are typically exchanged in batches of at most five per visit, mean circulation times are defined as the average elapsed time between needle distribution and return for returned needles normalized by the average number of needles distributed per batch, to obtain an average circulation time per needle. Such normalization ensures that the circulation times attributed to needles individually exchanged four times per week (i.e., in batches of size one) are equivalent to circulation times for needles exchanged weekly in batches of size four, for example.

Figure 1 shows the number of distinct client visits and the number of clients who visited the program each month since program inception through June 1993. That the decriminalization of syringe possession without prescription in Connecticut pharmacies apparently had a major impact on participation in the program is evident from this graph. If one focuses on the 20 months prior to July 1992, one notes that visitation frequency increased more rapidly than client participation, suggesting that the same clients were exchanging needles more frequently. This, in turn, suggests that needle circulation times should have fallen over time due to the increasing frequency of exchange.

Figure 2 documents the flow of needles in the New Haven program over time. Again the apparent impact of
Figure 1. Client participation and visitation.

Syringe decriminalization in July 1992 is visible. Also visible is the close match between the number of needles distributed and returned over time, indicating that for the New Haven program, the "law of conservation of needles" is approximately correct. It is clear that there is some leakage in the system. Some of this can be attributed to program policy whereby newly enrolling IDUs with no needles to exchange can receive a single "starter" needle, following which all exchanges are, in theory, to occur on a one-for-one basis. However, the amount of leakage that actually occurred exceeds what can be explained by this policy, indicating that in some transactions clients received more needles than they returned as a result of clerical error or deliberate deception.

Our theory postulates that more frequent exchanging should lead to a reduction in mean needle circulation times. Focusing only on the 20 months from November 1990 through June 1992 inclusive, Figure 3 shows that mean circulation times have been reduced. Thus, needle exchange appears to be interrupting the needle circulation process in the manner intended.

Figure 2. Volume of needle exchange.

Equation (12) affords another empirical test of the physics of needle exchange. Figure 4 plots $\xi(\nu)$ (defined monthly as the reciprocal of the mean circulation time) versus the distribution rate $\nu$. The strength of the linear relationship is clear from the figure. Regression yields the empirical relation

$$\hat{\xi}(\nu) = 0.04255 + 0.03492 \nu \quad (r^2 = 94.1\%) \quad (16)$$

(0.01992) (0.00206) (standard errors in parentheses).

Of great interest is the estimated value of $\hat{\alpha} = 0.04255$ for an estimate of preneedle exchange circulation times is provided by $\hat{\alpha}(0) = 1/\hat{\alpha} = 23.5$ days. As of the end of July 1992, circulation times had been reduced to between two and three days, so the reduction in needle circulation times appears to be substantial.

Equation (11) proposes an expression for the mean level of infection in circulating needles as a function of mean circulation times. Using the circulation time and PCR test results for the first 20 months of the program, we can estimate the unknown parameters $\pi_0$, $\lambda$, and $\mu$.  

Figure 3. Mean needle circulation time.

Figure 4. Needle exchange and removal rates.
via maximum likelihood. Specifically, letting $n_{\text{neg}}(i)$ denote the number of needles that tested positive for HIV via PCR in month $i$, $n_{\text{neg}}(0)$ denote the number of needles that tested negative in month $i$, and $\hat{p}(i)$ equal (11) with mean circulation time equal to $\bar{\gamma}$, the observed mean circulation time for month $i$), the likelihood function constructed is given by

$$L = \prod_{i=0}^{10} [\hat{p}(i)]^{n_{\text{pos}}(i)} [1 - \hat{p}(i)]^{n_{\text{neg}}(i)}.$$  

(17)

This function was maximized numerically using the GAUSS maximum likelihood routine (GAUSS 1992). The resulting maximum likelihood estimates are given by $\hat{p}_0 = 0.0016$ (standard error = 0.0025), $\lambda = 0.3675$ (standard error = 0.0676), and $\mu = 0.1665$ (standard error = 0.0824).

A number of results are worth mentioning. First, as all tests were performed on needle exchange needles one would expect that $p_0 = 0$. This intuition is confirmed statistically (the maximum likelihood estimate is not significantly different from 0). Second, the fit of the model can be gauged from Figure 5, which plots both observed and modeled mean infection levels in the needles over time. There are four clear outliers that cannot be explained by the theory, but the general fit of the model appears reasonably good.

Third, it is possible to conduct two "holdout" tests of the model. At the beginning of the needle exchange, 160 "street needles" delivered to the program in exchange for new needles were selected for HIV testing via PCR. Of these, 108 (or 67.5%) tested HIV positive (Heimer, Kaplan and Cadman 1992, Heimer et al. 1993). Although the true preneedle exchange circulation times are unknown, (16) affords the previously mentioned estimate of 23.5 days. Substituting this into (11) with the estimated parameters cited above yields an estimate of $\hat{p}(0) = 0.6375$ (standard error = 0.0529) incorporating fluctuations due to the sample size). That the prediction obtained is in such close accord with the level of infection in the street needles, needles which were not employed in estimating the parameters of (11), provides a certain degree of confidence in the model.

The second holdout test involves needles retrieved from an underground needle exchange program that existed prior to the legal program. Of 180 needles tested from the underground program, 113 (or 62.8%) tested HIV positive via PCR (Heimer, Kaplan and Cadman 1992, Heimer et al. 1993), extremely close to the prediction of (11). It is again important to point out that the needles from the underground program were not used to calibrate the model. A negative implication of this result, however, is that it appears that the underground program did not achieve a sufficient turnaround of needles to significantly change HIV transmission patterns. This example should again help to convince the reader that the results of these studies are not preordained. Negative results are possible, and occasionally they are obtained as in the example of the underground program.

For the legal needle exchange program, it appears that the impact of needle exchange on interrupting the needle circulation process, and hence the HIV transmission process, has been considerable. The mean circulation time of needles appears to have been reduced to near 3 days, so (13) suggests a relative reduction in HIV transmission on the order of 33% (obtained by setting $r = 23.5$ days and $\bar{\gamma} = 0.6375$). While this sounds impressive, recall that this is only a relative reduction in incidence via needle borne infections among needle exchange participants. Translating this relative reduction into an actual number of infections averted requires producing a range of baseline incidence estimates.

Consider first the equilibrium method based on (14). Prevalence surveys among New Haven drug injectors are not common, but a range of estimates are available. Data from the CDC national seroprevalence surveys included drug injectors sampled at a sexually transmitted disease clinic in New Haven, of whom 56.0% tested HIV positive (CDC 1990). On the high end, 67% of African American men entering a drug treatment program tested HIV positive (Dr. Patrick O'Conner, Yale University, personal communication), while early attempts to estimate HIV prevalence among needle exchange participants using more assumption driven models yielded estimates in the neighborhood of 60% (Kaplan and Heimer 1992). To be comprehensive, baseline HIV prevalence among needle exchange clients will be assumed to lie somewhere between 30% and 70% inclusive.

A number of studies have documented that the mean (or median) time between HIV infection and AIDS is approximately 10 years in duration (Bocchetti and Moss 1989, Brookmeyer and Goedert 1989, Freund and Book 1990, Munoz et al. 1989). Therefore, the mean progression rate $\gamma$ will be set equal to 0.10 per infected IDU per year. Equation (14) then suggests that aggregate HIV incidence at baseline among IDUs was somewhere between three and seven infections per 100 IDUs per year.

![Figure 5. HIV infection in needles over time.](image-url)
However, not all of these infections would have been transmitted via needle sharing. For now, it will be assumed that 30% of all infections among IDUs are acquired sexually, reducing the baseline incidence of needle borne HIV infections to between 2.1 and 4.9 per 100 IDUs per year. The New Haven data suggest a relative reduction in incidence on the order of 33%. Thus, the number of infections prevented is estimated to fall between 0.7 and 1.6 per 100 needle exchange participants per year. Since roughly 200 IDUs participated in the needle exchange on a monthly basis over the time period considered, the absolute number of infections prevented is estimated to fall between 1.4 and 3.3 annually. If one assumed that 40% of infections were transmitted sexually instead of 30%, the estimated number of infections averted would fall between 1.2 and 2.8. Given all available information, this approach suggests that roughly speaking, the New Haven needle exchange prevented between 1 and 3 infections annually.

The second approach to calculating the number of infections averted relies on backcalculation combined with estimates of the number of IDUs in New Haven. The backcalculation model was implemented using the Weibull incubation density estimated by Brookmeyer and Goelet; the median incubation time for this function equals 10 years. Using seven years of AIDS incidence data for New Haven drug injectors yielded a mean HIV incidence rate of approximately 160 infections per year. Assuming that 30% of infections are sexually acquired leaves a baseline annual incidence rate of 112 needle borne infections among New Haven's IDUs. Early attempts to estimate the number of IDUs in New Haven yielded estimates in the vicinity of 2,300 (Kaplan and Soloshat 1993), while it is difficult to believe that more than 5% of New Haven's population of 129,000 actively inject drugs, which yields an upper bound of 6,450 for the number of active IDUs. Considering a range of 2,000 to 7,000 IDUs suggests that the average needle exchange client population of 200 represents between 2.86% and 10% of the population of all drug injectors. Multiplying these percentages by the estimated 112 infections acquired via needle sharing yields a baseline between 3.2 and 11.2 infections annually. Finally, applying the 33% relative reduction suggests that between 1.07 and 3.73 infections have been averted annually. If we instead assume that 40% of all infections derived from sexual exposure, the estimated number of infections averted would fall between 0.92 and 3.2 per year, a range very close to that obtained from the equilibrium model.

It is worth mentioning that the lifetime medical expenses alone associated with treating a new HIV infection are estimated to total $119,000 (Hellingr 1993). The annual cost of the New Haven program is approximately $150,000, so even if one assumes that 40% of all HIV infections among clients are transmitted sexually, the program still appears to be cost effective. This calculation is more conservative than it may appear, for there are additional costs beyond medical expenses imposed by HIV and AIDS, while there are additional program benefits such as placement in drug treatment programs that we have not considered.

3. PUBLIC HEALTH POLICY

To say that needle exchange is controversial is an understatement. For example, Robert Martinez, former President Bush’s “drug czar” stated bluntly that “Our gains against drug use have been hard-won, and this is no time to jeopardize them by instituting needle exchange programs... there is no conclusive evidence that exchange programs reduce the spread of AIDS” (Martinez 1992). The results of the New Haven evaluation studies have proven important in challenging such views.

The methods and results reported in this paper have been presented in varying levels of technical detail to public health decision makers in the United States and abroad on numerous occasions. Together with the results of other needle exchange studies (for recent reviews see GAO 1993 and Lurie and Reingold 1993), this work has played its part in shaping policy in major cities such as Baltimore, New York, and San Francisco. In Connecticut, needle exchange continues in New Haven, and has been expanded to include the cities of Bridgeport and Hartford. While the U.S. government still officially opposes needle exchange programs, and no federal funding for such services is permitted, perhaps this too will change in the years ahead.

ACKNOWLEDGMENT

This research was supported in part by grant ROI DA07676 from the National Institute on Drug Abuse, and by grant 20049 from the Robert Wood Johnson Foundation.

REFERENCES


