Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a tabbed divider. Given our digitization capabilities, we are sometimes unable to adequately scan such dividers. The title from the original document is indicated below.

Divider Title: **5**

INFORMATION BRIEF:
RECENT EFFECTS OF INCREASED DRUG OFFENSE SANCTIONS ON STATE PRISON POPULATIONS

Kenneth E. Carlson

Tammy Enos

Abt Associates, Inc.

Executive Summary

The 1980s were characterized by broad increases in the severity of sanctions imposed on criminal offenders in most States. From 1980 to 1990, the number of sentenced prisoners held in State prisons increased at an average annual rate of 9 percent. In most years prior to 1985, drug offenders accounted for fewer than 10 percent of all prison admissions, and could therefore hardly have been a major factor in the growth of prison population. In 1985 and 1986, however, drug offenders accounted for 13 percent and 16 percent, respectively, of prison commitments, suggesting that their role in driving population growth was beginning to change.

Measures of drug consumption in the late 1980s suggest that the number of actual incidents of drug sales and possession did not increase much and may have decreased. Increases in the number of drug arrests are therefore much more likely to reflect increased enforcement intensity than actual increases in the number of incidents.

From 1985 to 1989, drug **arrests** (including possession, sale, manufacture, importation and distribution) in the 26 States included in this analysis increased at an average annual rate of 12 percent, while arrests for index offenses (including murder, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson) increased less than half as fast (4 percent per year). From 1989 to 1990, arrests for drug abuse violations decreased by 12 percent, while arrests for index offenses continued to increase at about the same rate as previously (3.7 percent). Thus, one reason for increasing use of the prisons for those convicted of drug law violations has been this growth in arrests.

In 1985, the ratio of drug prison **commitments** to drug arrests ranged from approximately 1 percent to 8 percent. By 1990, every State for which data were available had increased the ratio of commitments to arrests, and half of the States showed prison commitment rates in excess of 9.8 percent. Several States had also increased commitment ratios for other offenses but seldom by as much as for drug crimes. Between 1985 and 1990, commitments for drug offenses increased 10 times more than commitments for nondrug offenses (332 percent versus 44 percent). Half of the total increase in new commitments was made up of drug offenders, and 45 percent of the total increase can be attributed to the increased sanctions for drug offenses. Thus, a second reason for increasing use of the prisons for drug law violators is that when arrested and convicted, drug law violators were increasingly likely to be sentenced to prison.

In 1985, drug offenders served **prison terms** with averages ranging from 9 months to 48 months. By 1990, 11 of the 17 States for which data were available had decreased the average time served in prison by drug offenders, sometimes by more than 20 percent. Eleven of the 17 States also showed decreases in time served by nondrug offenders. Decreases in the average time served reflect a combination of effects. First, it is reasonable to expect that even **before** the major emphasis on drug law enforcement, the most serious **drug** offenders were being sentenced to prison. Other things **being** equal, increased incarceration rates could be achieved only by imposing prison sentences on less serious offenders previously serving probation terms. One would expect these less serious offenders to receive shorter prison terms than more serious offenders who were incarcerated previously. Second, in some state prison systems accelerated release may be applied as a population control measure in response to crowding problems created by increased commitments.

Although the States studied here have been affected to varying degrees by the increased sanctions for drug offenses, none have been untouched. Over the period from 1985 to 1990, every State included in this analysis has shown increases in both the number of drug offenders sentenced to prison and of drug offenders actively serving prison terms. These increases exceed the corresponding rates of increase for nondrug offenses for all States. There is considerable evidence that increases in commitments have been partially offset by decreases in time served, and some evidence that some States have decreased their incarceration of violent and property offenders to make room available to house more drug prisoners.

Altogether, the 26 States included in this analysis increased their count of drug prisoners from 25,000 in 1985 to 94,000 in 1990, while prisoners serving time for all other offenses increased from 253,000 to 323,000. If trends in arrest, commitment, and time served rates for nondrug offenses had been applied to drug offenders, the increase in the number of drug prisoners would have been only about 9,800. **Thus over 59,500 of the total 69,300 increase in drug prisoners could not be attributed to general arrest or incarceration rate increases. In these States the number of all prisoners (drug and nondrug) rose by approximately 139,000 during the five-year period. Forty-three percent of this increase is explained by direct and indirect effects of the increased sanctions for drug offenses.**[1]

[1] Statistics from the United States Courts indicate that increased prison commitments and time served for drug offenses are contributing substantially to increases in the number of inmates housed by the U.S. Bureau of Prisons. Drug offenders alone provided an increase of 5,980 annual arrivals (68 percent of the total increase from 1985 to 1990) at the Federal Bureau of Prisons. The length of prison sentences imposed on Federal drug offenders increased from an average of 58 months in 1985 to 81 months in 1990. During the same period the average sentence for nondrug offenses decreased from 46 months to 37 months.

It is still too early to observe changes in the time actually served in prison, but data on prison inmates indicate that 83 percent of drug offenders admitted in 1989 were still in prison 12 months after entry compared to 76 percent of those admitted in 1985. Two-thirds (66 percent) of drug prisoners admitted in 1989 were still incarcerated 24 months after commitment, compared to 60 percent of those committed in 1985. As the new sentencing policies take full effect, the number of Federal drug prisoners will increase substantially

Introduction

For nearly 50 years after the introduction of the National Prisoner Statistics series in 1925, prison populations increased no faster than the general growth of the population of the United States. With the exceptions of the Depression and the Second World War, prisons housed about one inmate for every thousand in the general population until approximately 1970.

> Throughout the 1960s, the size of the prison population of the United States was either stable or declining. Then, in 1972-73, that population began to rise sharply. In five years it increased by almost 50 percent.... In many jurisdictions, corrections resembled a rush-hour highway; the system runs normally at or above its capacity, so any abnormal load creates unmanageable difficulties.[2]

Subsequent experience has shown prison crowding, rather than being an abnormal condition, appears to have become the chronic condition of most prison systems. The trend of increasing population has now persisted for more than two decades. In recent years, research attention has turned toward identifying the underlying causes which have driven this sustained growth, both because it is an interesting sociological question and for the more practical reason that correctional planning and budgeting depend on reasonably accurate short and intermediate projections of future prison populations.

Various hypotheses have been reviewed in the literature[3] without producing a conclusive consensus on the sources of prison population growth. While these need not be completely recited here, the most widely discussed hypotheses include:

- increases in the number or seriousness of crimes,
- higher priority given to "getting tough on crime" in general public opinion or in the attitudes and actions of political leaders,
- increases in the crime-prone (or arrest-prone or prison-bound) demographic segments of society,
- unemployment, and

---

[2] Michael Sherman and Gordon Hawkins, *Imprisonment in America: Choosing the Future*, (Chicago: University of Chicago Press, 1981), 7.

[3] For example, Franklin E. Zimring and Gordon Hawkins, *The Scale of Imprisonment*, (Chicago: University of Chicago Press, 1991).

• drugs.

These theories are not mutually exclusive. Many logically consistent combinations can be imagined. Increased crime--or the perception thereof--might motivate public support for more incarceration. Additional crimes might be committed to finance drug consumption or as part of the business strategy of drug wholesalers. Although researchers disagree on the theory or combination of theories with which to explain the growing use of imprisonment in the past twenty years, most would probably agree that no single cause provides a complete account of past trends, and most would include demography and more severe sentences as two of the principle factors.

Early in 1991, Patrick Langan, of the Bureau of Justice Statistics, published an analysis of factors influencing changes in State prison commitments between 1974 and 1986.[4] He concluded that

> No clear evidence was found that prosecutors were increasingly using mandatory prison sentencing laws, that judges were imposing longer prison sentences than previously, or that parole boards were making prisoners serve longer before their first release. Changes since 1973 in population demographics and in police-recorded crime and arrest rates were found to have only a modest impact on prison population growth. The war on drugs was found to have only a small impact despite increased drug arrest and imprisonment rates. One change found to have a major impact was the increased chance of a prison sentence after arrest for nearly every type of crime.

This paper analyzes more recent data bearing on one specific finding of the Langan paper: the effect on prison populations of changes in the level of sanctions imposed on drug offenders. Few crimes have shown more dramatic increases in sanction severity over the past few years than drug offenses. In this paper recent data from several States is examined in an attempt to quantify the effects of increased drug enforcement on prison commitments and populations.

In most years prior to 1985, drug offenders accounted for fewer than 10 percent of all prison admissions,[5] and could therefore hardly have been a major factor in the growth of prison populations, which by then had been an established trend for more than a decade. In 1985 and 1986, however, drug offenders accounted for 13 percent and 16 percent, respectively, of prison commitments,

---

[4] Patrick Langan, "America's Soaring Prison Population," *Science*, 251 (29 March 1991): 1572.

[5] *Ibid.*, 1571.

suggesting that their role in driving population growth was beginning to change.

Individual State departments of corrections have specifically identified increased sanctions for drug offenses as a major contributor to recent increases in their populations.

> The key factor in the increase in court admissions [in Illinois] is, without doubt, the war on drugs and the subsequent increase in the number of drug offenders being sent to prison. Approximately 47% of last year's [FY 1991] population growth is attributable solely to more drug offenders coming to prison.[6]
>
> [In Florida,] increases in crime, especially drug arrests, are closely associated with the rise in new prison commitments.... Admissions to prison for new drug offenses have risen an astonishing 1,611% over the past ten years.[7]
>
> The number of [California's] felon new admissions for drug offenses increased by 433.9% from 2,428 in 1984 to 12,962 in 1989. As a percentage of all felon new admissions, drug offenders increased from 13.9% in 1984 to 37.9% in 1989. Drug offenders accounted for most of the overall increase in new admissions from 1984 to 1989.[8]
>
> ... The number of drug offender PV-WNTs [parole violators with new terms] has increased tenfold, from 379 in 1984 to 3,884 in 1989.[9]
>
> The dramatic increase seen in [Pennsylvania] court commitments in 1989 is most likely the result of new (1988 & 1989) mandatory sentences and increased enforcement measures aimed at drug offenders.[10]

---

[6] Illinois Department of Corrections, *Adult Projections Update: May, 1991*, 8.

[7] Florida Department of Corrections, *Annual Report: 1988/1989*, 24-25.

[8] California Department of Corrections, *California Prisoners & Parolees*, 1989, 14.

[9] *Ibid.*, 21.

[10] Commonwealth of Pennsylvania Department of Corrections, *Annual Statistical Report 1988 & 1989*, 1.

> An analysis of projections of future populations prepared by 12 States using a common projection model found that under existing policies, the States will increase their prison populations by over 68 percent by 1994 [compared to 1989], an average annual growth rate of about 13 percent per year.... The primary reason for the dramatic increase in prison populations is the War on Drugs, which is not only increasing the number of prison admissions but is also increasing the rate of parole violations. The already disproportionate rate of Blacks and Hispanics being sent to prison will increase considerably, principally due to the War on Drugs.[11]

The most recent nationally comprehensive data on prisoners extend only through 1986. Levels of drug enforcement have changed significantly since then, and analyses of the national historical data leave open the question of the impact of changes in the last five years. Although national data do not shed light on recent trends, many States have published reports which are sufficiently detailed to allow individual estimates of changes in their correctional systems through 1989 or 1990. The States gathering these statistics follow differing classification and counting practices, so that data from one State may not be comparable with data from others. Generally, however, calculations within each State are internally consistent, so that case study analysis of individual States is possible. The estimates presented here summarize the results of 26 such State level analyses.

The Data

Data were collected on arrests, prison commitments, and prison populations. To the extent possible, the data cover each year from 1980 to 1990. A special tabulation of the number of drug arrests in each State (1980-1989) was provided by the Uniform Crime Reports section of the Federal Bureau of Investigation. The police departments covered by these data correspond exactly to those reporting arrests published in the annual *Uniform Crime Reports* (UCR). The reporting jurisdictions include approximately 85 percent of the total population of the United States. (Coverage varied from State to State and year to year.) Each reported number of arrests was adjusted to estimate the total number of arrests which would have been observed had all jurisdictions reported. Arrest data for 1990 were computed by Abt Associates from individual agency reports provided by the UCR program. The estimation method used for 1990 incorporates data from incomplete reports, which are not used in the FBI's publications. The total number of drug arrests in the U.S. estimated by this method is 5

---

[11] James Austin and Arron McVey, "The 1989 NCCD Prison Population Forecast: The Impact of the War on Drugs," *NCCD Focus* (December 1989): 1.

percent higher than the corresponding estimate published by the FBI.

For analysis of arrest data, offenses were grouped into two categories: drug offenses (which includes possession, sale, manufacture, importation, and distribution) and index offenses. Index offenses include most of the serious offenses for which prison sentences are imposed. Index offenses include: murder (0.7 percent of the 3,006,000 index arrests estimated in the United States in 1989), forcible rape (1.3 percent), robbery (5.5 percent), aggravated assault (15.3 percent), burglary (15.6 percent), larceny-theft (53.4 percent), motor vehicle theft (7.6 percent), and arson (0.6 percent). Taken together, drug and index offenses account for approximately 80 percent of admissions to State prisons. Other arrests were not considered because the imprisonment rate for other offenses is much lower than that for drug and index offenses.

At the time files were closed for analysis, 26 States had provided data on prison commitments in both 1985 and 1990. We asked State respondents to restrict commitment data to newly sentenced inmates received from court, as opposed to returning parole violators and persons entering prison under other circumstances, but in six States where these data were not available, we accepted total commitments.

Numbers of prisoners were also provided by 26 States (not necessarily the same as those providing commitment data). In general, populations refer to inmates present on the last day of the State's fiscal year, although in some instances average daily populations or counts on some other date in the year were provided. Commitment data refer to the 12 months preceding the date on which populations were measured. The data generally describe adult prisoners held in State custody and therefore exclude inmates of county jails, even if temporarily housed there awaiting space at a State facility. Computed estimates of total prisoners in each State are internally consistent from year to year, but because of these differences in accounting practices, commitment and inmate data cannot be directly compared among States.

The Model

This analysis follows an approach generally similar to Langan's, with modifications imposed by the more limited data available on recent drug offenders. Langan modeled prison commitments as the product of

Population (grouped in three races and two age groups)

x Crime rate (grouped in nine offense categories)

x Probability that a crime results in arrest x Imprisonment rate.

To test the effect of specific changes, he applied 1974 rates for each factor to 1986 data. Drug offenses, crime rates and arrest probabilities are unknown, since drug sales and possession are generally not reported to the police. Thus Langan's model for these offenses is simplified to three factors: population, arrest rate, and imprisonment rate.

In the analysis presented here, demographic effects are excluded, first because Langan found them to be relatively uninfluential, and second because over the five years which have elapsed since Langan's data were collected, the composition of the population has not changed greatly. We have, however, added two components not analyzed by Langan. First, Langan's data refer to the combined prison populations of all 50 States. Since our data came from individual State reports, rather than national surveys, we could not replicate this aggregation. In examining the data from individual States, it became apparent that different States followed different policies with regard to sanctions for drug offenses; many showed greatly increased volumes of offenders and substantially higher use of incarceration, while others showed little effect at all. An aggregate analysis would dilute the magnitude of population changes in the States whose drug sanction policies changed most, although it would probably come to similar qualitative conclusions.

Second, this paper considers both prison commitments and prison populations. Prison population and the number of prisoners committed are directly related via the average time prisoners serve. The form of the relationship depends on the dynamics of commitment. If commitments are a linear function of time (t), the relationship is approximated by a simple product:

(1) $\text{prisoners}_t = \text{commitments}_{[t-1,\ t]} \times \text{time served}$

where: prisoners_t = the number of prisoners at time t,

$\text{commitments}_{[t-1,\ t]}$ = the total number of new commitments from court over the year from t-1 to t, and

time served = the average number of years expected to be served by an arriving prisoner.

Solving this equation for time served yields

$$\text{time served} = \text{prisoners}/\text{commitments}.$$

Although equation (1) is exact when prison commitments increase linearly over time, our data indicate that prison commitments and populations have been increasing significantly faster than linearly. A more accurate estimate of time served, which takes into account the effect of increasing commitments,[12] is provided by:

$$\text{time} = \frac{-\ln(1 - qr/(1 + r))}{\ln(1 + r)}$$

where: time = average time served,
r = annual rate of change in commitments, and
q = $\text{prisoners}_t/\text{commitments}_{[t-1,\ t]}$.

Prison commitments in this model are expressed as the number of arrests times the commitment rate:

(2) $\text{Commitments}_{[t-1,\ t]} = \text{arrests}_{[t-1,\ t]} \times \text{commitment rate}_{[t-1,\ t]}$

where: $\text{Commitments}_{[t-1,\ t]}$ = number of new prison commitments in the year t-1 to t,

$\text{arrests}_{[t-1,\ t]}$ = number of arrests in the year t-1 to t, and

$\text{commitment rate}_{[t-1,\ t]}$ = number of new commitments per arrest during the year.

This again is an approximation, since the persons arrested in one year are not necessarily sentenced in that year, the crimes reported at the time of arrest need not correspond to those for which offenders are sentenced, and some offenders are sentenced for crimes not covered by arrest data.

Combining equations (1) and (2) yields

(3) prisoners = arrests x commitment rate x time served.

Some States were unable to provide sufficient data to estimate time served but provided data on total numbers of prisoners. In these States a simplified version of equation (3) was used:

(4) prisoners = arrests x incarceration rate

where the incarceration rate is the ratio of prison population (rather than commitments) to arrest.[13]

---

[12] See appendix for derivation.

[13] The following States were modeled using equation (3): Delaware, Florida, Hawaii, Kansas, Massachusetts, Michigan, Minnesota, Montana, North Carolina, New York, South Carolina, and

The impact of the increased sanctions for drug offenses is estimated by comparing the actual number of prisoners in 1990 with the number which would be predicted by this model using hypothetical values of arrest, commitment, and time served for drug offenders. The hypothetical values are based on the assumption that each parameter for drug offenders would change from its 1985 level by the same proportion as did the corresponding factor for nondrug offenders.

This assumption combines trends which are not specific to drug offenses, such as demographic change and generally higher arrest or imprisonment rates, in a single factor. Over the five years from 1985 to 1990, changes in the composition of the population would have had only a slight effect on imprisonment; demographic factors can plausibly be advanced to explain trends over decades, but they change too slowly to be a major contributor to the more rapid changes which have characterized drug sanctions.

For the types of offenses commonly reported to the police (UCR index offenses), arrest rates are customarily based on either the number of crimes committed or the number of offenses known to the police. Over the period from 1985 to 1990, the number of these offenses increased at an average annual rate of 3 percent, while arrests for the same offenses approximately followed, increasing at 3.5 percent annually.[14] For drug offenses it is impossible to compare arrests with crime incidents, since these offenses are seldom reported. Indirect measures of drug consumption suggest that the number of actual incidents of drug sales and possession during the period covered by this analysis did not increase much and may have decreased:

> Between 1985 and 1988, use of all drugs in the past year decreased among those aged 12-17, except for hallucinogens and heroin, which remained relatively stable.
>
> ... Between 1985 and 1988, the use of most drugs in the past year decreased among those aged 18-25.

Washington. Six additional systems were modeled by equation (4): Arkansas, Alabama, Connecticut, the District of Columbia, Maryland, and South Dakota.

[14] Federal Bureau of Investigation, *Crime in the United States 1990*, (Washington: Government Printing Office, 1991), 50, 180.

> ... Between 1985 and 1988, the percentage of those aged 26 and older using drugs in the past year decreased for all drugs, although not all the decreases were statistically significant.[15]

Increases in the number of drug arrests are therefore much more likely to reflect increased enforcement intensity than actual increases in the number of incidents.

Arrests

From 1985 to 1989, estimated drug arrests in the United States increased at an average annual rate of 12 percent, while arrests for index offenses increased one-third as fast (4 percent per year). From 1989 to 1990, estimated arrests for drug abuse violations decreased by 12 percent, but arrests for index offenses continued to increase at about the same rate as previously (3.7 percent).[16] Changes from 1985 to 1989 in the number of drug arrests were unevenly divided among the States (figure 1). The number actually decreased in 11 States, and 10 more had compound annual increases of 10 percent or less (table 1).

Prison commitments

In 1985, the ratio of drug prison commitments to drug arrests (table 2) ranged from under one percent (Kentucky and Minnesota) to 8 percent (Mississippi). By 1990, every State for which data were available had increased the ratio of commitments to arrests, and half of the States showed prison commitment rates in excess of 9.8 percent. Several States had also increased commitment ratios for other offenses, but seldom by as much as for drug crimes.[17]

In the 26 States for which drug commitment data were available in 1985 and 1990, commitments for drug offenses quadrupled in five years (table 3). During the same period, commitments for nondrug offenses increased by a total of 44 percent. Commitments for all offenses rose by 118,000 prisoners. Half of this total increase in new commitments (62,000) was made up of drug offenders, and 45 percent of the total increase (54,000) can be attributed to the increased sanctions for drug offenses; that is, if drug arrests and

---

[15] National Institute on Drug Abuse, *National Household Survey on Drug Abuse: Main Findings 1988*, (Washington: Government Printing Office, 1990), 15, 16.

[16] FBI, *Crime in the United States 1990*, 182.

[17] Prison commitment rates for crimes other than drug offenses decreased in Delaware, Georgia, Hawaii, Illinois, Indiana, and North Carolina.

commitment rates had only kept pace with the corresponding statistics for nondrug offenses, prisons in these States would have received 45 percent fewer total new commitments.

Time served in prison

In 1985, drug offenders served prison terms with averages (table 4) ranging from 9 months (Georgia) to 48 months (Nebraska). By 1990, 11 of the 17 States for which data were available had decreased the average time served in prison by drug offenders, sometimes by more than 20 percent. Eleven of the 17 States also showed decreases in time served by nondrug offenders.

Some of these systems were operating under externally imposed population caps, which had the effect of requiring that one release accompany each admission. In these systems, net prison population changes were effectively insulated from increased commitment rates by a corresponding decrease in time served. Increased sanctions for drug offenders in these States can only result in increased backlogs of sentenced offenders held in local jails awaiting prison space[18] or early releases of prisoners convicted of violent or property crimes.[19]

Decreases in the average time served reflect a combination of effects which cannot be clearly distinguished on the basis of these data. First, it is reasonable to expect that even before the major emphasis on drug law enforcement, the most serious drug offenders were being sentenced to prison. Other things being equal, increased incarceration rates could be achieved only by imposing prison sentences on less serious offenders previously serving probation terms. One would expect these less serious offenders to receive shorter prison terms than the average of the more serious offenders who were incarcerated previously. Second, in some systems accelerated release may be applied as a population control measure in response to crowding problems created by increased commitments. Third, these measures of time served in prison do not include time served in local jails either awaiting trial or after sentencing pending transfer to State facilities. Both of these may have increased due to increased caseloads.

---

[18] "[As of December 1989] in Louisiana there are approximately 4,000 [State] inmates housed in local jails.... In Tennessee over 2,800 inmates, and in Virginia over 2,000 inmates are so housed." Austin and McVey, *"1989 Prison Population Forecast,"* 4.

[19] "Florida ... has dramatically increased its use of 'administrative gain time' and 'provisional release credits.' These actions have shortened prison terms to the extent that over 3,000 inmates are now being released early *each month*." *Ibid.* (emphasis in original).

Prisoners

Altogether, the 26 States providing consistent data on prisoners for this analysis increased their count of drug prisoners from 25,000 to 94,000, while prisoners serving time for all other offenses increased from 253,000 to 323,000 (table 5). If trends in arrest, commitment, and time served rates for nondrug offenses had been applied to drug offenders, the increase in the number of drug prisoners would have been only about 9800. Thus 59,500 of the total 69,300 increase in drug prisoners could not be attributed to general arrest or incarceration rate increases. In these States the number of all prisoners (drug and nondrug) rose by approximately 139,000 during the five-year period. Forty-three percent of this increase is explained by direct and indirect effects of the increased sanctions for drug offenses.

In four States (Nebraska, North Carolina, Washington, and West Virginia) the absolute number of prisoners serving time for violent and property offenses decreased. In these States the number of drug prisoners increased by substantially more than the decreases in the number of violent and property offenders. If these transfers are interpreted as displacement, these States provide instances where increased drug sanctions not only increased total prison populations but also decreased the States' ability to use prison as a punishment for other offenses. In these States, applying the changes in nondrug sanction rates could overstate the effect of drug offenders on total prison population, since some of the effect of increased drug prison commitments may have been to decrease another part of the total prison population.

Apart from these four States, there is little evidence that drug prisoners were being substituted for other types of offenders; the States with largest percentage increases in drug prisoners tended also to have greater than average percentage increases in the number of nondrug prisoners. The greatest impacts of increased drug sanctions appear to have been felt in Connecticut, the District of Columbia, Georgia, Massachusetts, New York, and Texas, where over half of the total increases in prison population were due to increases in the number of drug prisoners beyond that which would have been predicted based on trends in arrest, sentencing and time served for other offenses.

Federal Prisoners

Federal offenses differ substantially from those under State jurisdiction. Most of the violent and property offenses which result in State prison sentences come under Federal jurisdiction only in exceptional circumstances, such as when they are committed on Federal lands or are directed against victims for whom a Federal interest exists. For example, robbery generally is not covered by Federal law, but robbery of FDIC-insured banks is a violation of the U.S. Code. Thus, the preceding analysis which was appropriate for State drug arrests cannot be applied directly to DEA arrests.

However, statistics from the United States Courts indicate that increased prison commitments and time served for drug offenses are contributing substantially to increases in the number of inmates housed by the U.S. Bureau of Prisons. The number of convictions in the Federal courts for drug offenses rose from 10,289 in 1985 to over 16,000 in 1990.[20] (During the same period, the number of nondrug convictions increased by approximately one percent per year.) Incarceration rates for convicted drug offenders increased from 77 percent to 86 percent, while the rates for other kinds of offenses increased from 42 percent to 48 percent. The net result was that drug offenders alone provided an increase of 5,980 annual arrivals (68 percent of the total increase from 1985 to 1990) at the Federal Bureau of Prisons.

The length of prison sentences imposed on Federal drug offenders increased from an average of 58 months in 1985 to 81 months in 1990. During the same period, the average sentence for nondrug offenses decreased from 46 months to 37 months. These changes in sentences reflect the effect of sentencing guidelines, which applied to most of the offenders sentenced in 1990. Under the laws applicable in 1985, offenders sentenced to more than one year were usually eligible for parole after serving one-third of their sentences. Parole is abolished by the sentencing guidelines, and offenders are expected to serve their entire terms, except for a possible 15 percent reduction, which may be earned for good behavior. Thus, the 40 percent increase in nominal imposed sentences for drug offenders can be expected to translate into a substantially greater increase in time actually served.

It is still too early to observe changes in the time actually served in prison, but data on prison inmates indicate that 83 percent of drug offenders admitted in 1989 were still in prison 12 months after entry, compared to 76 percent of those admitted in 1985.[21] Two-thirds (66 percent) of drug prisoners admitted in 1989 were still incarcerated 24 months after commitment, compared to 60 percent of those committed in 1985. As the new sentencing policies take full effect, the number of Federal drug prisoners will increase substantially.

---

[20] Bureau of Justice Statistics, *Federal Criminal Case Processing, 1980-89, with Preliminary Data for 1990*, (Washington: Government Printing Office, 1991), 9.

[21] Unpublished analysis by Abt Associates Inc.

Conclusions

Although the States studied here have been affected to varying degrees by the increased sanctions for drug offenses, none have been untouched. Over the period from 1985 to 1990, every State with sufficient data to support analysis has shown increases in both the number of drug offenders sentenced to prison and of drug offenders actively serving prison terms. These increases exceed the corresponding rates of increase for nondrug offenses for all States. There is considerable evidence that increases in commitments have been partially offset by decreases in time served, and some evidence that some States have decreased their incarceration of violent and property offenders to make room available to house more drug prisoners.

APPENDIX

Estimation of time served in prison

When prison intake increases exponentially, equation 1 underestimates the time served. As an initial case, assume that all prisoners serve identical prison terms of s years

Let P_t = prison population at time t, and
C_t = new commitments from court during the year t-1 to t.

Then $P_t = P_{t-1} + C_t - C_{t-s}$

Equivalently, $P_t = \sum_{1}^{t} C_i - \sum_{1}^{t-s} C_i$

$$= \sum_{t-s+1}^{t} C_i$$

If commitments are increasing exponentially, $C_t = C_0 (1 + r)^t$, where r is the growth rate per period, and

$$P_t = C_0 \sum_{t-s+1}^{t} (1 + r)^i$$

$$= C_0 \frac{(1 + r)^{t+1} - (1 + r)^{t-s+1}}{r}$$

$$\frac{P_t}{C_t} = \frac{[C_0 / r] [(1 + r)^{t+1} - (1 + r)^{t-s+1}]}{C_0 (1 + r)^t}$$

$$= \frac{(1 + r) - (1 + r)^{1-s}}{r}$$

$$= \frac{(1 + r) [1 - (1 + r)^{-s}]}{r}$$

Solving this equation for s yields

$$s = \frac{-\ln (1 - rP_t/C_t/(1 + r))}{\ln (1 + r)}$$

where: s = average time served and
r = annual rate of change in commitments

TABLE 1
DRUG AND OTHER ARRESTS, 1985 and 1990

State	Drug Offenses 1985	Drug Offenses 1990	Drug Offenses % Change	Model for 1990	Net Effect	Other Offenses 1985	Other Offenses 1990	Other Offenses % Change
Alabama	7,684	13,398	74.4%	8,686	4,712	30,171	34,105	13.0%
Alaska	1,756	758	-56.8%	1,081	-323	10,329	6,359	-38.4%
Arizona	12,833	14,471	12.8%	16,563	-2,092	48,281	62,311	29.1%
Arkansas	6,084	6,061	-.4%	6,910	-849	19,884	22,581	13.6%
California	179,723	251,605	40.0%	240,410	11,195	340,113	454,961	33.8%
Colorado	7,667	7,556	-1.4%	7,839	-283	47,868	48,941	2.2%
Connecticut	13,267	21,374	61.1%	16,395	4,979	44,170	54,585	23.6%
Delaware	1,334	2,222	66.6%	1,651	571	6,464	8,000	23.8%
District Of Columbia	9,146	9,998	9.3%	10,132	-134	10,374	11,492	10.8%
Florida	42,753	67,983	59.0%	57,547	10,436	158,247	213,010	34.6%
Georgia	19,221	33,326	73.4%	28,005	5,321	59,135	86,162	45.7%
Hawaii	5,167	3,597	-30.4%	6,609	-3,012	9,847	12,596	27.9%
Idaho	1,520	1,858	22.3%	1,837	21	9,850	11,907	20.9%
Illinois	35,881	80,456	124.2%	56,133	24,323	60,553	94,731	56.4%
Indiana	7,576	10,325	36.3%	9,935	390	38,903	51,015	31.1%
Iowa	2,761	3,311	19.9%	3,505	-194	20,471	25,989	27.0%
Kansas	3,605	6,334	75.7%	2,677	3,657	20,992	15,588	-25.7%
Kentucky	11,593	13,598	17.3%	11,447	2,151	30,706	30,318	-1.3%
Louisiana	10,500	17,049	62.4%	11,480	5,569	53,293	58,266	9.3%
Maine	1,759	2,329	32.4%	1,840	489	9,934	10,393	4.6%
Maryland	20,147	28,686	42.4%	22,953	5,733	53,969	61,487	13.9%
Massachusetts	16,005	22,403	40.0%	20,865	1,538	38,712	50,467	30.4%
Michigan	16,814	28,330	68.5%	19,841	8,489	77,530	91,489	18.0%
Minnesota	5,551	7,005	26.2%	6,019	986	33,959	36,823	8.4%
Mississippi	8,042	11,116	38.2%	8,517	2,599	32,530	34,450	5.9%
Missouri	11,466	17,462	52.3%	13,479	3,983	65,803	77,358	17.6%
Montana	1,086	1,626	49.7%	919	707	8,145	6,894	-15.4%
Nebraska	2,371	4,064	71.4%	2,583	1,481	13,452	14,657	9.0%
Nevada	5,840	7,838	34.2%	6,720	1,118	14,902	17,146	15.1%
New Hampshire	2,248	4,146	84.4%	2,149	1,997	6,822	6,520	-4.4%
New Jersey	36,920	47,225	27.9%	40,711	6,514	77,330	85,270	10.3%
New Mexico	5,914	6,405	8.3%	6,451	-46	20,595	22,465	9.1%
New York	103,544	124,591	20.3%	124,375	216	184,679	221,833	20.1%
North Carolina	17,389	26,504	52.4%	24,295	2,209	61,096	85,361	39.7%
North Dakota	952	939	-1.4%	1,160	-221	5,126	6,243	21.8%
Ohio	22,755	30,436	33.8%	26,620	3,816	85,541	100,067	17.0%
Oklahoma	9,084	8,947	-1.5%	9,383	-436	28,620	29,561	3.3%
Oregon	6,394	9,914	55.1%	6,739	3,175	33,671	35,491	5.4%
Pennsylvania	19,578	27,193	38.9%	15,567	11,626	94,296	74,981	-20.5%
Rhode Island	3,017	2,939	-2.6%	3,725	-786	8,013	9,895	23.5%
South Carolina	11,319	15,766	39.3%	14,684	1,082	26,350	34,185	29.7%
South Dakota	1,504	1,042	-30.7%	1,321	-279	6,999	6,147	-12.2%
Tennessee	10,768	23,457	117.8%	12,741	10,716	49,371	58,419	18.3%
Texas	61,550	62,634	1.8%	71,073	-8,439	180,369	208,277	15.5%
Utah	5,729	4,026	-29.7%	6,631	-2,605	25,163	29,125	15.7%
Vermont	741	883	19.2%	737	146	2,462	2,451	-.4%
Virginia	12,431	18,312	47.3%	17,074	1,238	45,495	62,487	37.3%
Washington	8,911	17,419	95.5%	9,138	8,281	60,592	62,131	2.5%
West Virginia	1,647	2,706	64.3%	1,581	1,125	11,873	11,392	-4.0%
Wisconsin	9,857	9,401	-4.6%	10,700	-1,299	58,226	63,201	8.5%
Wyoming	957	708	-26.0%	1,049	-341	3,393	3,719	9.6%
ALL OBSERVATIONS	**822,359**	**1,141,732**	**38.8%**	**1,010,481**	**131,251**	**2,444,668**	**2,923,300**	**19.6%**

Source: Uniform Crime Reports, Except 1990 drug arrests recomputed by Abt Associates.

TABLE 2

PRISON COMMITMENTS AS A FRACTION OF ARRESTS FOR DRUG AND OTHER OFFENSES, 1985 and 1990

State	Drug Offenses 1985	Drug Offenses 1990	Other Offenses 1985	Other Offenses 1990	Model for 1990	Net Effect
Alabama		11.8%				
Arizona	3.8%	13.5%	7.3%	9.6%	5.0%	8.5%
Arkansas		10.6%		11.9%		
California	2.0%	5.5%	5.0%	5.6%	2.3%	3.2%
Delaware	4.8%	11.2%	10.6%	7.5%	3.4%	7.7%
Florida	4.7%	23.8%	7.8%	13.4%	8.1%	15.7%
Georgia	5.1%	10.6%	10.1%	8.2%	4.1%	6.5%
Hawaii	1.1%	3.2%	4.3%	3.8%	1.0%	2.2%
Idaho	5.1%	13.8%	8.7%	10.6%	6.2%	7.6%
Illinois	2.2%	5.1%	15.3%	13.5%	1.9%	3.2%
Indiana	6.7%	9.6%	13.0%	10.1%	5.2%	4.4%
Iowa		7.5%		4.7%		
Kansas	5.5%	10.8%	8.0%	13.8%	9.4%	1.3%
Kentucky	.9%	1.4%	3.2%	4.0%	1.1%	.3%
Massachusetts	2.2%		5.3%			
Michigan	3.0%		9.3%			
Minnesota	.8%	3.6%	3.8%	4.5%	.9%	2.7%
Mississippi	8.0%	10.1%	8.2%	9.2%	9.0%	1.1%
Missouri	2.9%	7.6%	6.8%	9.6%	4.1%	3.5%
Montana	2.9%	5.0%	3.1%	3.5%	3.3%	1.7%
Nebraska	1.7%	9.9%	4.8%	6.1%	2.2%	7.7%
New Hampshire	1.4%	2.8%	4.3%	8.4%	2.8%	.1%
New York	2.1%	8.7%	5.5%	5.6%	2.2%	6.5%
North Carolina	5.6%	9.7%	16.3%	12.2%	4.2%	5.5%
Ohio	5.4%	16.4%	10.2%	12.4%	6.6%	9.8%
Oklahoma	5.0%	17.6%	12.7%	15.2%	6.0%	11.6%
Pennsylvania	1.8%		4.2%			
South Carolina	7.4%	14.6%	22.5%	25.7%	8.4%	6.1%
Tennessee		10.0%		7.1%		
Texas	3.9%	14.4%	8.4%	9.4%	4.4%	10.0%
Vermont		8.3%				
Virginia	2.8%	15.0%	7.6%	11.0%	4.1%	10.9%
Washington	1.2%	6.7%	2.5%	2.8%	1.4%	5.3%
Wisconsin	1.9%	6.4%	4.5%	4.9%	2.1%	4.4%
COMPLETE OBSERVATIONS*	**3.7%**	**9.9%**	**8.2%**	**9.3%**		

* Unweighted averages

Source: Abt Associates Inc., 1991.

TABLE 3
PRISON COMMITMENTS FOR DRUG AND OTHER OFFENSES, 1985 and 1990

State	Drug Offenses		Other Offenses		Model	
	1985	1990	1985	1990	for 1990	Net Effect
Arizona	484	1,952	3,517	6,006	827	1,125
Arkansas		640		2,696		
California	3,609	13,741	16,895	25,502	5,448	8,293
Delaware	64	248	682	602	56	192
Florida	2,013	16,169	12,380	28,595	4,650	11,519
Georgia	976	3,529	5,946	7,049	1,157	2,372
Hawaii	57	115	426	477	64	51
Idaho	78	257	857	1,258	114	143
Illinois	787	4,094	9,271	12,815	1,088	3,006
Indiana	505	987	5,048	5,135	514	473
Iowa		248		1,213		
Kansas	197	681	1,686	2,159	252	429
Kentucky	103	194	974	1,207	128	66
Massachusetts	359		2,050			
Michigan	503		7,186			
Minnesota	42	249	1,278	1,651	54	195
Mississippi	640	1,127	2,656	3,179	766	361
Missouri	329	1,328	4,452	7,410	548	780
Montana	32	81	253	238	30	51
Nebraska	41	402	641	896	57	345
New Hampshire	32	118	295	547	59	59
New York	2,218	10,785	10,202	12,313	2,677	8,108
North Carolina	982	2,582	9,955	10,383	1,024	1,558
Ohio	1,240	5,000	8,760	12,409	1,757	3,243
Oklahoma	453	1,572	3,624	4,498	562	1,010
Pennsylvania	346		3,960			
South Carolina	834	2,294	5,916	8,801	1,241	1,053
Tennessee		2,355		4,160		
Texas	2,417	9,048	15,117	19,600	3,134	5,914
Vermont		73				
Virginia	351	2,755	3,442	6,892	703	2,052
Washington	109	1,165	1,486	1,719	126	1,039
Wisconsin	188	606	2,647	3,125	222	384
COMPLETE OBSERVATIONS	**18,781**	**81,079**	**128,406**	**184,466**	**27,257**	**53,822**

Source: Abt Associates Inc., 1991.

TABLE 4

ESTIMATED TIME SERVED IN PRISON (YEARS) FOR DRUG AND OTHER OFFENSES, 1985 and 1990

State	Drug Offenses 1985	Drug Offenses 1990	Drug Offenses % Change	Model for 1990	Net Effect	Other Offenses 1985	Other Offenses 1990	Other Offenses % Change
Arkansas		1.35				2.07		
California	1.60	2.10	31.0%	1.75	.34	2.78	3.05	9.6%
Delaware	1.80	1.79	-.3%	2.54	-.74	2.22	3.12	41.1%
Florida	1.31	.57	-56.2%	.65	-.08	2.32	1.16	-50.1%
Georgia	.74	.67	-9.3%	.70	-.03	1.84	1.74	-5.1%
Hawaii	2.08	1.29	-38.1%	2.12	-.83	3.08	3.15	2.1%
Illinois	1.01							
Iowa		1.41					2.94	
Kansas	1.37	1.10	-19.6%	1.23	-.12	2.64	2.36	-10.6%
Massachusetts	.88					2.26		
Michigan	1.64					2.35		
Minnesota	1.64	1.14	-30.6%	1.53	-.39	1.90	1.77	-6.9%
Mississippi	1.09	1.31	19.6%	1.10	.21	2.22	2.23	.5%
Missouri	1.92	1.38	-27.9%	1.60	-.22	2.23	1.86	-16.5%
Montana	1.98	2.11	6.8%	2.61	-.50	3.86	5.10	32.0%
Nebraska	3.96	4.97	25.4%	1.66	3.32	2.72	1.14	-58.2%
New York	2.66	1.97	-25.9%	2.69	-.72	3.04	3.08	1.2%
North Carolina	1.34	1.12	-16.6%	1.25	-.13	1.61	1.49	-7.2%
Oklahoma	.81	1.38	69.0%	.80	.57	2.32	2.29	-1.4%
Pennsylvania	1.77					3.45		
South Carolina	1.10	1.42	29.3%	1.08	.34	1.64	1.61	-1.6%
Tennessee		.84					2.70	
Texas	1.16	1.02	-12.1%	1.03	-.01	2.37	2.10	-11.5%
Vermont		.62						
Washington	1.97	1.11	-43.7%	1.68	-.58	4.40	3.76	-14.4%
COMPLETE OBSERVATIONS	**1.38**	**1.30**	**-4.6%**	**1.31**	**-.01**	**2.39**	**2.33**	**-4.3%**

17 states dist. avg. [illegible]

Source: Abt Associates Inc., 1991.

TABLE 5

PRISONERS INCARCERATED FOR DRUG AND OTHER OFFENSES, 1985 and 1990

State	Drug Offenses			Other Offenses			Combined Model for 1990	Net Effect
	1985	1990	% Change	1985	1990	% Change		
Alabama	901	2,26	151.5%	10,040	13,134	30.8%	1,179	1,087
Alaska	105	187	78.1%	2,019	2,240	10.9%	116	71
California	5,355	25,136	369.4%	43,769	71,629	63.7%	9,552	15,584
Connecticut	550	2,661	383.8%	5,148	6,859	33.2%	733	1,928
Delaware	104	402	286.5%	1,534	1,932	25.9%	143	259
District Of Columbia	1,672	4,304	157.4%	4,798	5,303	10.5%	1,848	2,456
Florida	2,480	10,074	306.2%	25,830	32,659	26.4%	3,032	7,042
Georgia	744	2,462	230.9%	10,771	12,142	12.7%	811	1,651
Hawaii	110	145	31.8%	1,283	1,466	14.3%	135	10
Kansas	259	743	186.9%	4,279	4,934	15.3%	310	433
Massachusetts	316	1,502	375.3%	4,624	6,051	30.9%	414	1,088
Maryland	575	2,159	275.5%	11,906	13,883	16.6%	670	1,489
Minnesota	62	277	346.8%	2,377	2,867	20.6%	83	194
Mississippi	695	1,447	108.2%	5,758	6,927	20.3%	841	606
Missouri	560	1,746	211.8%	9,332	13,200	41.4%	877	869
Montana	58	155	167.2%	995	1,245	25.1%	79	76
Nebraska	124	1,400	1029.0%	1,647	1,013	-38.5%	95	1,305
New York	4,655	18,459	296.5%	29,852	36,436	22.1%	7,210	11,249
North Carolina	1,278	2,860	123.8%	15,980	15,476	-3.2%	1,277	1,583
Oklahoma	377	2,068	448.5%	8,188	10,023	22.4%	451	1,617
South Carolina	907	3,126	244.7%	9,443	13,838	46.5%	1,341	1,785
South Dakota	48	115	139.6%	921	1,223	32.8%	64	51
Texas	2,744	9,194	235.1%	34,576	39,963	15.6%	3,213	5,981
Vermont	7	45	542.9%	613	955	55.8%	11	34
Washington	175	1,259	619.4%	6,222	6,218	-.1%	212	1,047
West Virginia	83	94	13.3%	1,581	1,471	-7.0%	77	17
ALL OBSERVATIONS	**28,075**	**97,926**		**293,856**	**350,426**			
COMPLETE OBSERVATIONS	**24,944**	**94,286**	**278.0%**	**253,486**	**323,087**	**27.5%**	**34,772***	**59,514***

Source: Abt Associates Inc., 1991.

* Detail may not add to total due to rounding.

Clinton Presidential Records
Digital Records Marker

This is not a presidential record. This is used as an administrative marker by the William J. Clinton Presidential Library Staff.

This marker identifies the place of a tabbed divider. Given our digitization capabilities, we are sometimes unable to adequately scan such dividers. The title from the original document is indicated below.

Divider Title: **6**

A PRELIMINARY PLAN
FOR ESTIMATING THE NUMBER OF "HARD-CORE" DRUG USERS

Ronald Simeone
William Rhodes
Dana Hunt

ABT ASSOCIATES, INC.

Revised February 12, 1993

This proposal has been submitted to ONDCP for internal review only. The proposal has not received technical review or full editing.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This proposal, prepared at the request of the Office for National Drug Control Policy, discusses procedures for estimating the number of "hard-core" drug users in selected cities. These estimates are used in conjunction with more readily available social indicators--such as the numbers of drug related arrests and deaths due to narcotism--to develop synthetic estimates (SE) of hard-core drug user in geographic areas where no direct estimates exist.

The estimates require a mathematical model. The parameters for that model would be estimated using multivariate statistical analysis of field surveys and institutional contact data. The former would be derived from probability samples (specifically, the National Household Survey on Drug Abuse) as well as opportunistic samples (interviewing heavy users where they can be found). The latter would be collected from administrative records on arrests, hospitalizations, and drug treatment admissions.

The mathematical model has derivative benefits with far reaching implications for drug policy research. It will provide an empirical profile of drug use, and drug-related arrest, hospitalization, and treatment. It will identify when events of each type typically occur in the careers of heavy users, as well as how long they last. The empirical model will provide the basis for a simulation of drug use and institutional contact. With this model the impact of a potential change in national policy could be assessed in advance of its implementation. For example we might attempt to determine how drug use and help-seeking behavior vary as a function of drug treatment availability. Analyses such as these could be instrumental in projecting demand for criminal justice, health, and drug treatment services.

Our plan has four phases. During Phase I, we will develop a model of drug use and institutional contact activity based on extant data. That preliminary model will serve as a prototype for a revised model to be based on new data. We will also use this preliminary model to make estimates of the number of hard-core drug users in the United States. During Phase II, we will develop survey instruments and data collection procedures in collaboration with the relevant agencies in at least one city. These instruments and procedures will be specifically designed to capture the population of heavy users who are under-represented in existing surveys. Phases III and IV will involve a pilot project to collect data in at least one city and to establish a national network of reporting sites, respectively. We would expect to receive funding from the Office of National Drug Control Policy (ONDCP) for the first and second phases. The third and fourth phases would require additional support from either the National Institute on Drug Abuse (NIDA) or Substance Abuse and Mental Health Services Agency (SAMHSA).

INTRODUCTION

Major policy initiatives have been pursued by the Federal government in an effort to reduce the number of drug users in the United States. These have involved ambitious programs focusing on both supply and demand reduction. While these efforts are believed to have been effective, it is difficult to measure their success empirically due to an absence of reliable information on the size of the drug using population. Although large-scale surveys have received broad application in the area of public health research and are known to provide accurate prevalence and incidence measures for a wide range of behavioral phenomena, a consensus has formed that self-report information collected via both in-person and telephone interviews consistently underestimates the number of "hard-core" drug users. This phenomenon is reflected at both the State and national level.[1]

These limitations are to be expected. Many "street addicts" are simply not reachable using conventional sampling methods because they don't have telephones or permanent homes; thus, accessibility to this population is an issue. Further, drug use remains a highly stigmatized behavior, and the criminal justice sanctions associated with both possession and sale of controlled substances can be severe. Reticence often becomes the rule when respondents answer questions about drug use; thus the veracity of self-report information is suspect.

In response to these apparent weaknesses in survey methods, a number of researchers have attempted to use data from other sources in deriving estimates of the number of hard-core drug users. This has involved the application of mathematical models that are believed to be reasonable approximations of the processes by which people who use drugs make contact with various elements of the criminal justice, health, and drug treatment systems. These models are predicated on the belief that observed "institutional contacts" can tell us something about unobserved drug use activity. If this is indeed the case, we can avoid many of the shortcomings associated with the use of survey-based estimates. *Local estimates could be made of the number*

[1] Telephone surveys conducted in New York, for example, routinely produce estimates for the number of recent heroin users that are one-quarter the size of the State's outpatient methadone maintenance population. Similarly, general population surveys conducted by the National Institute on Drug Abuse (NIDA) using in-person interviews have reported declines in drug use even as other indicators more attuned to changes in the size of the hard-core population of drug users held steady or suggested a countervailing trend.

of hard-core drug users, based upon a combination of administrative record and survey information. These could serve as the "anchor points" to be used in the development of synthetic estimates for various levels of geographic aggregation, including the Nation as a whole.

Two conditions must obtain for these estimates to be valid: *it must be possible to model the behavior under study--hard-core drug use--with a reasonable degree of accuracy; and sufficient data must exist at the local level to support such estimates.* We examine these conditions in detail in the sections that follow.

CONVENTIONAL APPLICATIONS

Researchers have applied two general classes of techniques in the analysis of administrative record data for the purpose of estimating the size of the hard-core drug using population. These include *capture-recapture* and *truncated Poisson* models. Although some material has been written concerning the use of each technique (Frank et al., 1978; Sheridan, Levine, and Ney, 1983; Woodward, Bonett, and Brecht, 1985; Brecht and Wickens, 1992; Hser et al., 1992; Rhodes, 1993; Simeone, Nottingham, and Holland, 1993), little systematic research examines the underlying assumptions of these models and the extent to which they are consistent with the phenomena under study.

In support of this proposal, we completed a comprehensive evaluation of the applicability of capture-recapture and truncated Poisson models to the drug use estimation problem. Our detailed findings are provided in Appendix I of this document. When reading this material it is important to bear in mind that our analysis focused upon conventional applications of capture-recapture and truncated Poisson models, meaning applications in which administrative records constitute the sole source of information used in the estimation procedure. Such information, which can be collected only *after* an individual is arrested, hospitalized, or is admitted to treatment, must include a unique case identifier and dates of arrival and departure.

Our assessment of administrative record-based applications is quite negative. But it is necessary that we review them in detail since it is this critique that provides guidance to the development of an alternative, and methodologically sound, strategy for adapting existing mathematical models to the drug use estimation problem.

For the purpose of exposition, we have limited the following discussion to an evaluation of administrative record based capture-recapture analyses, and we will take as our example the process by which people are admitted to and released from drug treatment. Capture-recapture models are often applied to the drug use estimation problem in this manner, and the material that will be presented in subsequent sections is largely an extension of these techniques. Our conclusions, however, can be generalized to other methods and to other types of institutional contact.

Capture-Recapture Models

Biologists originally developed capture-recapture models to estimate the size of animal populations. In the model's most basic form, it is assumed that the system under study is closed, which is to say that no births, deaths, in-migrations, or out-migrations occur. If populations are relatively stable in size and we are not interested in examining population dynamics over time, then this assumption is generally tenable. When such techniques are applied to a population of animals under study, lake trout for example, individual members are typically netted, marked, and released during a first sample. Some time is then allowed to elapse so that the marked fish are able to disperse randomly within the unmarked population, and the representation of these marked individuals is then examined in a second netting. Thus:

$$\hat{N} = M\left(\frac{n}{m}\right) \qquad [1]$$

where $\hat{N}$ is the estimated size of the population, M is the number of marked members in the population, n is the number of individuals in the net, and m is the number of marked individuals in the net. Much work has been done by way of developing more sophisticated closed system capture-recapture models, but the essence of the technique remains unchanged (see, for example, Bishop, Feinberg, and Holland, 1975; Bonett, Woodward, and Bentler, 1986). A number of critical assumptions are inherent in even this most simple model: that marks are indelible; that encounter probabilities are homogeneous; that the delay in captivity is both brief and inconsequential in terms of affecting subsequent encounter probabilities; and, perhaps most significantly, that the ratio of marked to total members in the net is identical to the ratio of marked to total members in the non-captive population.

When applied to the problem at hand, this approach has typically involved identifying individuals admitted to and released from treatment during a given period of time as "marked," and as described above, allowing some period of time to elapse before determining the number of "recaptured" individuals in a second identically defined sample (Frank et al., 1978; Woodward, Bonnett, and Brecht, 1985).

A number of variations on this simple theme exist which were designed to examine the dynamic behavior of populations over time. Jolly (1965) and Seber (1965, 1973) developed one of the most broadly cited of these "open-system" models. The Jolly-Seber model involves repeated sampling of the population at successive time intervals. During each capture period, unmarked individuals are "tagged" and released, along with other previously marked individuals. Within this context, additional parameters are estimated for rates of birth and death and for in-migration and out-migration. Otherwise, the assumptions underlying this open system model are essentially the same as those associated with the closed system model discussed earlier. Although it has often been presented as a potentially useful technique, the only published application of this method to the problem under examination here involved the use of simulated data on drug use and institutional contact (Woodward, Bonnett, and Brecht, 1985; Brecht and Wickens, 1992).

In some ways the application of administrative record-based capture-recapture models to the problem of drug use estimation has great intuitive appeal, but the enterprise presents a number of unique challenges, which ultimately prove to be insurmountable.

The first of these is that it is difficult to form an operational definition of the population being estimated when capture-recapture techniques are applied in this manner. This is due to a violation of the mark indelibility assumption that is inherent in both the closed and open system capture-recapture models. It comes down to this: A lake trout that is captured, marked, and released remains a lake trout until the day it dies. An addict, on the other hand, does not remain in a constant state of drug use following treatment and discharge. He or she may resume use immediately, may stop permanently, or may resume use following some period of abstention. Nor do individuals remain in a constant state of drug use *prior* to capture, for that matter. Drug use is, in fact, episodic, and technically "marked" individuals have *zero* probability of entering treatment, or being recaptured, so long as they refrain from drug use. This problem cannot be regarded in terms of out-migration as described in the by traditional open system model, since individuals who are treated, released, and temporarily abstain may effectively migrate back into the hard-core population but not conform to the traditional understanding of an in-migrant--a new member of the population. Thus, individual's migration into and out of the identity of hard-core user makes defining M (the number of "marked" individuals) in Equation [1] very difficult. A second problem has to do with the extent to which a legitimate analogy can be drawn between animal capture and release and specific kinds of institutional contact. Capture-recapture models assume that the duration of delay in the captive state is relatively brief. Obviously some disjuncture occurs when we attempt to equate captivity with any form of institutional contact, such as drug treatment, that introduces a substantial delay into the process. Individuals must be admitted and released during one sample interval and admitted and released again during a second sample interval in order to be "marked" and "recaptured", respectively. This would not constitute a violation of model assumptions, were it not for the fact that people differ in their propensity to seek and complete treatment. This means that the most active "treatment-seekers" will be over-represented in the "net" when capture-recapture methods are implemented in the manner described above, and any estimate of N will be negatively biased as a consequence.

A third problem results from a fundamental distinction that must be made between help-seeking behavior, which is volitional, and capture, which requires only passivity on the part of an animal under study. This distinction has profound consequences in terms of limiting the applicability of record based capture-recapture methods to the task at hand. The issue here is that certain forms of institutional contact, such as admission to drug treatment, require purposeful action on the part of individuals. This means that, in the absence of the difficulties described above, it is only possible to estimate the size of the "treatment-prone" population--not the size of the drug using population in its entirety--when capture-recapture models are applied in the conventional way. Advocates of administrative record based capture-recapture models often refer to this as treatment "susceptibility". More critical reviewers have simply characterized the estimates as ill-defined.

A fourth and final problem arises from the fact that it is impossible to examine changes in the size of the drug using population when capture-recapture models are applied in this manner. Once again, this is particularly true when treatment admission and release data are used in the estimation procedure. The difficulty here is that it often takes years before a drug user chooses to seek treatment for his addiction, and substantial delays are also likely to exist between first use of a particular drug and initial contact with various components of the criminal justice and health care systems. Thus, models that make use only of information collected at the point of institutional contact are of very limited utility when examining the dynamics of drug use.

ALTERNATIVE APPLICATIONS

The preceding section identified a number of problems in the use of an administrative record based capture recapture model, and in this section we reformulate the capture-recapture model to more accurately reflect drug use behavior over time. As we shall see, it is not that capture-recapture models are intrinsically inconsistent with the process under study, but rather that they may not be applied using administrative records alone.

A Basic Reformulation

We return now to examine the essential characteristics of Equation [1], the closed system capture-recapture model. There are two basic components of the model: a quantity of M (marked individuals in the population), which is known, and a ratio of marked to total individuals in the sample, m/n, which is also known. It is not possible to estimate N from these terms using only administrative records for the reasons described in the preceding section. However, it is possible to formulate an analog to the capture-recapture model with terms that can be estimated using a combination of administrative record and survey data.

For the moment, let us proceed under the assumption that we are able to estimate, using survey data, the ratio of the number of drug users who are active and who seek treatment during some period, to the total number of drug users who are active during this same period. We denote the population parameter for this ratio as Q and our estimate as $\hat{Q}$. Similarly, let us assume that we are able to collect information, from administrative records, on the number of individuals who are admitted during the period referenced in the survey. If we denote this quantity as M, then:

$$\hat{N} = \frac{M}{\hat{Q}} \quad [2]$$

Equation [2] is simply a representation of the closed system capture-recapture model. The estimated ratio of "marked" to "total" elements, $\hat{Q}$, is derived from survey rather than from administrative record data. The admission data are used exclusively to provide scale. Although we will not elaborate on the point at this juncture, it is clear that if information on M were collected for a successive number of intervals, and if $\hat{Q}$ were estimated for these same intervals, a corresponding reformulation of the open system capture-recapture model would be possible. We note at this point that each of the four problems identified in the preceding section is solved when this estimation approach is adopted.

Specifically, survey data can be used to estimate the parameters of a model of drug use behavior. That model can tell us the percentage of users who never enter treatment, overcoming the problem that such users cannot be identified in treatment contact data alone. That model can also tell us the rate at which treated users relapse and again seek treatment, overcoming the problem that treated users differ from untreated ones. Furthermore, the model can tell us about periods of remission, overcoming the problem that treated users are inactive sometimes for short, and sometimes for long, periods of time. We elaborate on this model later.

Defining the Population

Of course, this solution begs the question: what is a representative sample of hard-core drug users?" Our analysis now requires an operational definition for hard-core drug users, as well as a method for knowing something of the composition of this group.

Several alternatives exist for operationalizing this definition, but we limit our discussion here to two which reflect the most common tendencies in drug use research. These include the measurement of level of use and the measurement of dysfunction. The level of drug use is often characterized by recency and frequency. It is customary to create categories such as "regular use" and "heavy use" based on informed judgements about what comprises especially problematic use. If it is drug using behavior per se that is the object of study, then this mode of measurement is adequate. If, on the other hand, it is drug using behavior that is believed to be related to some specific pathology which is the object of study, then alternative measurement strategies are more appropriate. One of these involves the use of DIS (Diagnostic Interview Schedule) operationalizations of DSM-III-R criteria. For the present, we note only that either approach may be suitable given the task at hand.

The problem of "representativeness" poses more daunting challenges. As defined here, q is simply the ratio of individuals who report a given form of contact, m, to the total number of individuals, n (the sample). Thus:

$$q = \frac{m}{n} \qquad [3]$$

As a practical matter, q is difficult to estimate using survey data. Because probability-based surveys fail to reach many hard-core drug users, we must supplement probability-based samples with opportunistic samples. Of course, this raises a new problem. We are uncertain that opportunistic samples are truly representative of users whose drug use careers we want to model. Furthermore, we must confront the reality that many hard-core drug users will deny their use, and thus, that useful interviews can only be conducted with current users who are willing to admit their use. This means that those hard-core drug users who are temporarily abstaining must be excluded from the sample, and this omission introduces a host of biases. We discuss how we will deal with those biases in the following section.

Estimation of Q from Survey Data

In order to deal with the problems mentioned above, we introduce the concept of a drug use career. A drug use career comprises various types of events, each of specific duration. These events may include active drug use, treatment, arrest, hospitalization, and abstinence. For our purposes, we will speak from a simplified model that includes only episodes of first "heavy" drug use and subsequent periods of treatment, recovery, and relapse.[2]

[2]As mentioned earlier, "heavy" use of drugs can be defined in different ways. The specific definition need not concern us here.

A career thus begins the first time a person uses drugs at or above some operational threshold. Thereafter, heavy drug use runs in spells of relapse and recovery. Let the function $U_i(t) = 1$ if the i^{th} person uses drugs heavily at time t. Let $U_i(t) = 0$ when he or she does not. A graph of $U_i(t)$ would show $U_i(t)$ at zero before the initiation of a drug use career, at 1 for some time after initiation, and at 1 or 0 intermittently throughout that career.

For simplicity we assume that drug treatment always precedes periods of recovery and that periods of relapse always precede drug treatment.[3] We represent the probability of entering drug treatment as $Tr_i(t)$ at time t. Thus, the i^{th} user initiates his drug use at time T_{i1}^b and ends it temporarily at time T_{i1}^e. He starts heavy drug use again at time T_{i2}^b and ends it temporarily at time T_{i2}^e, and so on. Treatment occurs at times T_{i1}^e, T_{i2}^e, and so forth. Now define U(t) as:

$$U(t) = \frac{\sum_{i=1}^{I} U_i(t)}{I} \qquad [4]$$

and define Tr(t) as:

$$Tr(t_2 - t_1) = \frac{\sum_{i=1}^{I} \sum_{\tau=t_1}^{\tau=t_2} Tr_i(\tau)}{I} \qquad [5]$$

U(t) is the probability that an individual will be using heavily at time t, and $Tr(t_2-t_1)$ is the probability that an individual will enter treatment between time t_1 and t_2.[4] In fact, we would

---

[3] In reality, drug users frequently stop using drugs heavily without entering treatment. Also, some users enter and remain in treatment without stopping the use of drugs. It is unnecessary to complicate the exposition with these considerations. We also note that other events (arrests, episodes of emergency room treatments, and so on) are part of the drug use career. It is convenient but not essential that we restrict the discussion to drug abuse treatment.

[4] We have assumed that the probability of more than one treatment admission between t_1 and t_2 is zero.

want to model $U(t)$ and $Tr(t_2-t_1)$ as mathematical representations of the career observed in the data. Appendix IV provides details.

If we knew $Tr(t)$ for a representative sample of drug users, we could estimate Q:

$$\hat{Q} = \frac{Tr(t_2-t_1)}{U(t)} \qquad [6]$$

Q is just the estimated probability that the average user's career would include an episode of treatment between time t_1 and t_2. By assumption this is also the span of time for which we have a treatment admission sample. With Q and M, we can estimate N, the total number of users with active drug use careers. Then the total number of people who are current users (in other words, those not in recovery) can be estimated as:

$$total\ active\ users = N \cdot U(t)$$

The concept of a drug use career has several advantages. Some of these will be discussed in some detail in a subsequent section. Here, however, we turn to the question of whether we can actually describe the characteristics of a drug use career given the data at our disposal.

It would be a relatively simple matter to estimate the parameters of drug use careers for users in an ordinary probability sample. We would begin by adopting a tentative mathematical model that represented the occurrence and duration of episodes of heavy drug use, treatment, recovery, and relapse. Findings presented by Simeone, Nottingham, and Holland (1993) might provide guidance in this effort. We have observed, for example, that in New York State the time between first use of heroin and first admission to treatment follows a lognormal distribution. Subsequent treatment admissions follow what appears to be a contagious Poisson process, in which the elapsed time between release and readmission diminishes as a function of the number of prior admission episodes. Taken together the lognormal and contagious Poisson distributions provide a model of the probability that a drug user will enter treatment during a given interval of time. We would next apply a statistical methodology appropriate for analyzing data on repeated events and might find that the convoluted lognormal-contagious Poisson model did not fit the data as well as we expected. Consequently, we would explore other mathematical representations of the process and continue the process of "fitting" until a suitable model was found. However, this illustration assumes that we have a probability sample. In fact, the sample planned for this study can not have a probability basis. Rather, we will interview heavy users where they congregate. (Appendix III provides a discussion.) Can the sample planned for this study be applied as easily? We have to contend with two serious problems: over-representing the most accessible user and dealing with under-reporting.

Over-representation of the most accessible users

Our sample may over-represent one type of user and, consequently, under-represent another type. For example, it may be easier to find and interview older males than it is to find and interview younger female users. Our measure of Q, based on the drug use careers of these predominately older, male users, may misrepresent the careers of other users.

The first step when dealing with this problem is to sample users from multiple sources. The sample must include all types of users. Appendix III discusses our strategy for accomplishing this goal. A second step requires us to slightly modify the form of Equation [2]. Although it is not possible to know much about the size or composition of the population of hard-core drug users on an a priori basis, it is possible to identify subpopulations whose members have relatively homogeneous probabilities of entering treatment. If sex and age (younger than 21 years of age versus 21 years of age and older), for example, were two relevant factors, then we would denote these estimated probabilities as $\hat{Q}_{j,k}$, j,k = 1,2. For our purposes it is necessary only that:

$$\hat{Q}_{j,k} = \frac{M_{j,k}}{N_{j,k}} \tag{7}$$

and therefore, that:

$$\hat{N}_{j,k} = \frac{M_{j,k}}{\hat{Q}_{j,k}} \qquad [8]$$

where $M_{j,k}$ is the number of admissions occurring during a given interval for members of a particular subpopulation, and $\hat{N}_{j,k}$ is the estimated size of the subpopulation. The total number of hard-core drug users is given by summing over all categories in the matrix. Thus, assuming that j and k are subscripts corresponding to sex and age, respectively:

$$\hat{N} = \sum_{j=1}^{2} \sum_{k=1}^{2} \hat{N}_{j,k} \qquad [9]$$

It is important to note at this point that sex and age are being used here purely for the purpose of exposition and that any number of factors might be used to partition the sample. What is essential is that a sample be constructed that allows reasonable estimates for the elements of $Q_{j,k}$ to be made, and thus the notion of "representativeness" pertains only to these elements. Once we have established a representative sample, we must contend with the accuracy of the self-reported information hard-core drug users provide.

Under-reporting by drug users

A large percentage of drug users deny their use. Ideally we would survey both those who currently use drugs heavily and those who are in recovery. We would ask them to report, retrospectively, sufficient information to provide an empirical profile of their careers. Unfortunately this approach does not work when a large percentage of users deny their current drug use. We also surmise that current abstainers will be harder to locate than current users, because abstainers may abandon drug using associates and locales. Either way, the empirical profile will provide a biased representation of drug use careers.

Our approach is to survey those individuals who admit to current use of drugs. This approach overcomes some problems while it introduces others. We can partially overcome the problem of denial of drug use by interpreting an admission of current use as indication that a survey respondent will be truthful when reporting his or her retrospective drug use history. We will then apply a calendar technique to improve the accuracy of these responses. Appendix III discusses the calendar approach.

However, this approach introduces a new problem into the analysis. By sampling only current users we obviously under-represent periods of abstinence. This problem is known as truncation. We can overcome this problem, by adjusting the statistical model to accommodate the truncation. Basically, the approach is to frame a mathematical model of behavior, and then determine how the sample statistics must be adjusted account for this truncation. We do not present the mathematics of that adjustment here, but we sketch an approach in Appendix IV.

Our plans dictate that we ask survey respondents to report retrospectively about the timing of periods of drug use, drug treatment, arrests, and use of emergency rooms. Knowledge of the timing of these events is essential to developing the profile of the drug use career. In turn, we can estimate Q from this profile, and eventually, the number of hard-core users.

A Simple Extension

To this point we have explained how prevalence estimates can be derived from a combination of survey data and institutional records. Although prevalence estimates are important, they may not be the most important aspect of this study. More important are the implications that knowledge of the dynamics of the drug use career has for the development and implementation of public policy.

The function $U_i(t)$ represents only one aspect of a drug use career: whether someone is or is not using drugs at time t. Similarly, the function $Tr_i(t)$ represents just one other aspect of a drug use career: whether or not a drug user enters treatment at time t. Other functions are required to fully represent the drug use career. For instance, drug users are sometimes employed and sometimes unemployed; they actively commit nondrug crimes at certain times and are law abiding at other times; they are incarcerated at some times and at liberty during other periods. These and other aspects of the drug use career would be included in our comprehensive model.

A comprehensive model is a neat way of conceptualizing and answering some questions that are fundamental for drug abuse policy research. The most obvious issue is determining the number of active heavy users at a given time. However, other questions are equally useful to drug control policy: How frequently do heavy drug users come into contact with the public health system? If emergency rooms and medical clinics are frequently visited by heavy users, these may be points of intervention for prevention programs. How frequently do heavy users come into contact with the criminal justice system? If this contact is as frequent as it appears to be, then prevention programs delivered through the criminal justice system may prove most effective. (The criminal justice response need not be punitive: programs such as TASC often serve as a powerful incentive for hard-core users to seek treatment.) Thus, the model describes the interaction between drug users and the various systems that provide social control and services.

Treatment effectiveness is integral to our model. Following treatment, drug use and criminal activity decline, and employment increases, at least for some users and for some length of time. These changes have cost-benefit implications. Thus, our model provides a framework for evaluation of treatment effectiveness.

Our model could be used to understand the effectiveness of other supply-based interventions as well. How responsive are drug users to the price of drugs and alcohol? How responsive are drug users to the use of intensive local enforcement (street sweeps and other law enforcement interventions? How responsive are they to prevention programs?

Summary and Conclusion

We have developed a method that successfully modifies capture-recapture models to estimate changes in the hard-core drug using population. It requires survey information to estimate the probability that members of various drug using subpopulations will have contact with relevant institutions during a given interval. By using survey rather than administrative record data, we avoid bias resulting from the volitional nature of capture. The approach also has a distinct advantage over conventional applications in its ability to detect changes in the size of the drug using population. Since the probability of entering into treatment is not calculated using administrative records, the delays that exist between incidence and institutional contact do not serve to preclude the examination of dynamic behavior. Further, the model can provide an invaluable framework for understanding what we do and do not know about the drug use careers of heavy drug users. ***Our method appears to provide a suitable method for satisfying the first*** *condition of our study's viability: we can model the drug use career with data that are likely to become available.*

SYNTHETIC ESTIMATION

The preceding sections describe a model which may be used to derive estimates of the size of the hard-core drug using population. We have recommended an approach based on the premise that we can establish local data bases which contain both survey and institutional contact data. We can then use these data bases to produce local estimates that will become the "anchor points" in a more general exercise in synthetic estimation of drug use at the national level.

Estimation Considerations

Synthetic estimation is a fairly straightforward technique which has received broad application, particularly in the field of mental health. It is based on the premise that if the relationship between relevant indicator variables and the anchor points chosen can be measured with a reasonable degree of accuracy, then the size of the population may be inferred in areas where reliable local indicators alone are available. In practice, anchor point information constitutes the dependent variable in a regression equation where the independent variables are a set of more readily available indicators (see Rhodes, 1993 for a detailed discussion and Appendix II of this proposal for a review of recent examples). Ideally, these indicators are selected on the basis of some theoretical justification, although it is also possible to identify suitable variables largely on the basis of their predictive utility.

Let us assume, for example, that we have derived estimates of N using the reformulated closed system capture-recapture model that was discussed in the preceding section, and that $\hat{N}_i$ constitutes the estimated number of hard-core drug users in each of 20 cities. If we believe that N_i is a linear function of some independent variable, perhaps the number of emergency room mentions involving drugs, here denoted as x_i then:

$$N_i = \beta_0 + \beta_1 x_i + e_i \qquad [12]$$

where β_0 is an intercept term, β_1 is a slope, or "weight" and e_i indicates simply that the relationship between $\hat{N}_i$ and x_i is measured with error. Therefore, we may define the total population of hard-core drug users in the 20 cities, here denoted as T, exclusively in terms of β_0, β_1, and the observed values for x_i. Thus:

$$T = \sum_{i=1}^{20} (\beta_0 + \beta_1 x_i) \qquad [13]$$

where e_i is assumed to be normally distributed with a mean of 0.

If we can be relatively certain that these coefficients remain constant from place to place, then we can also estimate the size of the hard-core drug using population in areas where no information on Q is available. This is so because $\hat{N}$ is defined as a function of x. By extension, it would be possible to estimate the number of hard core drug users in, for example, the largest 100 cities in the United States, using only 20 capture-recapture based estimates of $\hat{N}$ and 100 observations of x. Obviously, the accuracy of our estimates would be increased by incorporating additional explanatory variables into Equation [12]. Suitable candidates variables might include the numbers of drug-related arrests, emergency room mentions of heroin and cocaine, deaths due to narcotism, and gonorrhea and syphilis cases.

Summary and Conclusion

The preceding sections have elaborated upon a reformulated capture-recapture model, that draws upon administrative records and survey information in developing local estimates of the size of the hard-core drug using population. In this section, we have demonstrate that these anchor points may be expressed as a function of various social indicators in a process of synthetic estimation.

DATA REQUIREMENTS

There are two sets of information requirements that must be considered at this juncture: those related to the construction of anchor points, and those related to the process of synthetic estimation. These are discussed in detail below.

Anchor Points

Information collected through the use of a general population survey alone would lead to biased estimates of Q. This is in part because some individuals cannot be reached using conventional survey methods (the problem of accessibility) and in part because some individuals who are surveyed do not admit to drug use (the problem of veracity).

It is possible to get better estimates of Q by supplementing information collected during the course of a general population survey with information collected using more unconventional means. This would require interviewing large samples of individuals found in known "copping areas", of shelter inhabitants, of "couch people", and so forth. A discussion of various issues involved in collecting such data is provided in Appendix III.

The manner in which we have defined Q, and the strategy that we describe above for supplementing general population survey data with information derived from other samples, directly address the problem of accessibility. We must now address the problem of veracity. Our argument here is that it would be possible to at least partially verify survey responses if unique case identifiers--both for interviewees and for individuals who have contact with the relevant institutions--were maintained in a consistent manner. Comparison of survey and administrative record data would allow us to identify under-reporting of drug use and institutional contact and to improve estimates of $\hat{N}$ accordingly.[5] However, as a practical matter, it may be quite difficult to gain access to administrative records for this purpose. Treatment providers are understandably reluctant to identify their clients.

---

[5]In the best of circumstances, we can imagine establishing an integrated data base, maintained perhaps at the city level, that incorporates a unique case identifier which remains consistent across each component of the health, criminal justice, and drug treatment systems. If this were the case, we could simultaneously assess the accuracy of information provided by respondents as well as the reliability of our estimates of $\hat{N}$. We note at this point that while a data base of this kind would be very desirable, it is not essential to completion of the proposed work.

An alternative, and perhaps more realistic, method for dealing with the problem of veracity may be to interview only individuals who admit to current drug use. We reason that a person who is willing to be candid concerning his or her current use of an illicit substance is likely to be honest regarding other elements of his or her drug use career. This solution introduces a problem: this model under-represents periods of abstinence. Fortunately, we can correct for this bias using a mathematical model described in Appendix IV.

Synthetic Estimation

Synthetic estimation can be seen as a problem in regression analysis. The dependent variable is the estimate of the number of heavy drug users in sites for which survey and administrative data make estimation possible using the techniques already discussed in this proposal. The independent variables are indicators of drug use activity. These might include emergency room mentions of drug use, medical examiner reports, arrests for drug possession crimes, and so on. The regressions would be estimated using data from sites for which the dependent variable is known. The regression would then be used to predict the prevalence of drug use in sites where only the independent variables are known.

The accuracy of these predictions depends on several factors, not all of which are known at this time. If the relationship between the dependent and independent variables is very strong, then accurate predictions could be derived from a regression that is calibrated from just a few observations. For example, if the ratio of heavy cocaine users to emergency room mentions is consistently between 15 and 17, then predictions based on just a few anchor points would be adequate. If, in contrast, the ratio was between 10 and 30, we might require many more anchor points to be confident in predictions based on emergency room mentions of cocaine use.

The problem with having a large number of anchor points is that the cost of developing estimates increases roughly linearly with the number of anchor points. Having more than, say, 10 or 20 anchor points may be prohibitively expensive.

This observation suggests that while we may continue to think of our data collection efforts as being conducted in selected cities, we should derive the estimates for any given city from data collected at a lower geographic level--the neighborhood, or known heavy drug use areas. This would substantially increase the number of anchor points available for analysis.

This approach only requires that location information be collected for survey respondents and that indicator data be available at this same level geographic location.

As a practical matter, it is possible to collect information on social indicators at the zip code, or even the census tract, level in some cases. There are some idiosyncracies of reporting here that would have to be dealt with, and it is possible that they would vary to some extent from city to city. However, we can probably expect that drug related arrest data would be maintained only at the police precinct level and that health indicators might be subject to aggregation only at the city health planning district area. The drug users themselves may pose a more difficult problem. Since many hard-core users are transient, it may be difficult to associate them with a given residence or even location, at least so far as self-report address information is concerned. Suffice it to say that during the process of interviewing, some more novel method may have to be employed in order to derive proxies for geographic area data where necessary. These are issues that must be examined within the context of developing a detailed implementation plan, Phase II of our proposal.

Summary and Conclusion

The preceding discussion reviews some of the basic information requirements associated with the construction of anchor points and the process of synthetic estimation, respectively. ***Our analysis leads us to conclude that the data requirements associated with adopting the proposed model are entirely reasonable, satisfying the second condition of our study's viability: we are able to collect sufficient data at the local level to support our study.***

PHASES OF THE STUDY

The material presented thus far has provided a basic theoretical model for estimating the size of the hard-core drug using population in selected areas of the country and has identified the general data requirements associated with using this model in a process of synthetic estimation. In implementing this study, we envision four developmental phases.

Phase I

During the project's first phase. Abt staff members will use extant data to develop the mathematical model that will be used for further estimation. We think that preliminary parameter estimates can be derived from the National Aids Demonstration Research Project (NADR) and the National Household Survey on Drug Abuse. We have extensive experience with both sources. Other sources (such as the Drug Abuse Forecasting Project) might provide some confirmation.

During this phase of the project we would also develop some preliminary estimates of the size of the heavy user population. Two surveys -- DSRS and NDATUS -- could then be used to derive some estimates of the number of users who were in treatment in at a given time. (We also have extensive experience with these two data files.) Then, using the modified capture-recapture methodology outlined in this proposal, we could develop national estimates of the number of heavy drug users.

We would not have great faith in the resulting estimates. The NADR date were not collected with this exercise in mind, and they are hardly well designed for the contemplated analysis. The DSRS and NDATUS, similarly, have serious deficiencies when applied to this task. Nevertheless, using these and possibly other data files will be extremely instructive as we refine a methodology for estimating the parameters of the drug use career.

Preliminary estimation would also allow us to test the sensitivity of our mathematical model to the assumptions that underlie its derivation. Using this model we would develop a simulated system of drug use, treatment, arrests, and health care. We would simulate results under different model configurations to determine the model's sensitivity to assumptions made.

This simulation exercise would serve another purpose. It is difficult in the abstract to determine what information should be collected from users during interviews. Preliminary work on the simulation model should provide important insight into what we absolutely have to know (and thus, subject areas that should get the most attention in our interviews), and areas that are less crucial. Thus, this work should help guide the development of a survey instrument.

Phase II

Simultaneous with Phase I. we will produce a detailed blueprint for implementing a demonstration project in a specific site. perhaps Boston or New York City. This blueprint would focus principally upon issues related to data requirements. as described in the preceding section. and would result in the design of survey instruments, as well as the specification of data collection procedures. We also expect to field test the instruments during this phase. In addition. we would explore the feasibility of using a unique case identifier in this research and the extent to which the health. criminal justice. and drug treatment agencies in various cities are receptive to this approach. Our plan would thus be shaped, to some extent, by a number of concerns related to feasibility. As part of this process. we would establish appropriate agreements with organizations in one city to participate in the demonstration.

Phases III and IV

In the third phase of the project we would implement and evaluate a demonstration project; in the fourth phase, we would establish a national network of sites for collecting institutional contact information as well as survey data for members of "hard to reach" populations. These efforts would be coordinated with existing federally supported survey activities. Because the third and fourth phases of the project are likely to be multi-year initiatives, we believe that it would be appropriate to attempt to secure funding from NIDA or SAMHSA for them after successfully completing Phases I and II.

Personnel

Dr. Ronald Simeone will manage this task. In this capacity he will be responsible for project administration and for the estimation of drug use prevalence. Dr. William Rhodes will collaborate with Dr. Simeone on all phases of the project and will be responsible for developing mathematical models of drug use careers. Dr. Dana Hunt will also collaborate with Drs. Simeone and Rhodes and will have responsibility for designing and supervising all data collection activities.

REFERENCES

Amemiya, T. 1984. Tobit models: A survey. *Journal of Econometrics* 24: 3-63.

Bishop, Y.M.M., Feinberg, S.E., and Holland, P.W. 1975. *Discrete multivariate analysis: theory and practice*. Cambridge, MA: MIT Press.

Bonett, D.G., Woodward, J.A., and Bentler, P.M. 1986. A linear model for estimating the size of a closed population. *British Journal of Mathematical and Statistical Psychology* 39: 28-40.

Brecht, M. L., and Wickins, T.D. A comparison of multiple-recapture models for drug prevalence estimation.

Dhrymes, P. 1984. Limited dependent variables. In Griliches, Z., and Intriligator, M. (eds). *Handbook of econometrics*. Vol. 2. Amsterdam: North-Holland.

Frank, B., Schmeidler, J., Johnson B., and Lipton, D.S. 1978. Seeking truth in heroin indicators: The case of New York City. *Drug and Alcohol Dependence* 3: 345-358.

Hser, Y., Anglin, D., Wickins, T., and Homer, J. 1992. *Techniques for the estimation of illicit drug-use prevalence: an overview of relevant issues*. Washington, D.C.: National Institute of Justice. NCJ Research Report No. 133786.

Jolly, G.M. 1965. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* 52: 225-247.

Maddala, G. 1983. *Limited dependent and qualitative variables in econometrics*. New York: Cambridge University Press.

Maddala, G. 1984. Disequilibrium, self selection, and switching models. In Griliches, Z., and Intriligator, M. (eds). *Handbook of Econometrics*. Vol. 2. Amsterdam: North-Holland.

Rhodes, W. 1993. Synthetic estimation applied to the prevalence of drug use. *Journal of Drug Issues*. (Forthcoming).

Seber, G.A.F. 1965. A note on the multiple-recapture census. *Biometrika* 52: 249-259.

Seber, G.A.F. 1973. *The estimation of animal abundance and related parameters*. London: Griffin.

Sheridan, J.R., Levine, G.L., and Ney, N.S. 1983. *1982 Maryland study of drug abuse incidence, prevalence, and treatment demand*. Silver Spring, Maryland: Maryland Drug Abuse Administration. Contract No. 52524.

Simeone, R.S., Nottingham, W.T., and Holland, L. 1993. Estimating the size of a heroin using population: An examination of the use of treatment admissions data. *International Journal of the Addictions* 78 (Forthcoming).

Woodward, J.A., Bonett, D.G., and Brecht, M.L. 1985. Estimating the size of a heroin-abusing population using multiple-recapture census. Rockville, Maryland: National Institute on Drug Abuse. Research Monograph Series No. 57, pages 158-170.

APPENDIX I

AN EVALUATION OF ADMISSION-BASED CAPTURE-RECAPTURE AND TRUNCATED POISSON MODELS

CONVENTIONAL APPLICATIONS OF EXISTING MODELS

Two general classes of techniques have been used to analyze treatment admission data to estimate the size of the hard-core drug using population: *capture-recapture* and *truncated Poisson* models. Although some documentation of the use of each technique exists (Frank et al., 1978; Sheridan, Levine, and Ney, 1983; Woodward, Bonett, and Brecht, 1985; Brecht and Wickens, 1992; Hser et al., 1992; Rhodes, 1993; Simeone, Nottingham, and Holland, 1993), little systematic research has been conducted that examines the underlying assumptions of these models and the extent to which these assumptions are consistent with the phenomena studied.

In this appendix, we review several variations of capture-recapture and truncated Poisson models and discuss their potential for application to the estimation of hard-core drug use. Dr. Simeone has conducted preliminary research using these techniques in a computer model that simulated individual drug using and help seeking behaviors.[1] We used the results of this study to inform the conclusions and proposals that follow, and while we may present them in a deductive fashion, in fact, many insights are the product of manipulation and experimentation.

When reading this section, it is important to bear in mind that our analysis focuses upon the evaluation of conventional applications of capture-recapture and truncated Poisson models. By conventional, we mean that admission data constitute the sole source of information used in the estimation procedure. This information, which can be collected only *after* an individual presents for treatment, typically includes dates of arrival and departure and dates of first use for various substances. Our assessment of such admission-based applications is quite negative, but it is important that we review them in detail since it is this critique that provides guidance in the development of an alternative, and methodologically sound, strategy for adapting capture-recapture and truncated Poisson models to the drug use estimation problem.

---

[1] See Simeone, Nottingham, and Holland (1993). The authors developed a FORTRAN based individual level simulation routine, which is essentially an operationalization of the system depicted in Figure 1. The program accommodates the use of variable parameters for rates of entering drug treatment and durations of stay in drug treatment, and generates individual records showing the timing of various events. It is thus possible to derive estimates of the number of drug users within this simulated system by applying various capture-recapture and truncated Poisson models to the data base which is produced. The authors' delay functions used a gamma distribution generator developed by Phillips (1971) and provided in Shannon (1975).

Capture-Recapture Models

Capture-recapture models were developed in the biological sciences for the purpose of estimating the size of animal populations. In their most basic form, an assumption is made that the system under study is closed, which is to say that no births, deaths, in-migrations, or out-migrations occur. If populations are relatively stable in size and we are uninterested in examining population dynamics over time, then this assumption is generally tenable. When such techniques are applied to a population of animals under study, lake trout in this example, individual members are typically netted, marked, and released during a first sampling. Some time is then allowed to elapse so that the marked fish are able to disperse randomly within the unmarked population, and the representation of these marked individuals is then examined in a second netting. Thus:

$$\hat{N} = M\left(\frac{n}{m}\right) \qquad [1]$$

where $\hat{N}$ is the estimated size of the population, M is the number of marked members in the population, n is the number of individuals in the net, and m is the number of marked individuals in the net. Much work has been done by way of developing more sophisticated closed system capture-recapture models, but the essence of the technique remains unchanged (see for example Bishop, Feinberg, and Holland, 1975; Bonett, Woodward, and Bentler, 1986). A number of critical assumptions are inherent in even this most simple model: that marks are indelible; that encounter probabilities are homogeneous; that the delay in captivity is both brief and inconsequential in terms of affecting subsequent encounter probabilities; and perhaps most significantly, that the ratio of marked to unmarked members in the net is identical to the ratio of marked to unmarked members in the non-captive population.

As Woodward, Bonnett, and Brecht (1985) have observed, the consequences of violating these assumptions have been well examined (Carothers, 1973; Gilbert, 1973) and are known to include negative bias in the case of population heterogeneity. Within this context, variability in encounter probability can be accommodated by homogeneous grouping but only to the extent that such variability is not attributable to capture per se.

When applied to the problem at hand, this approach normally involves identifying individuals admitted to and released from treatment during a given period of time as "marked," and as described above, allowing some period of time to elapse before determining the number of "recaptured" individuals in a second identically defined sample (Frank et al., 1978; Woodward, Bonnett, and Brecht, 1985).

A number of variations on this simple theme exist which are intended specifically for the purpose of examining the dynamic behavior of populations over time. One of the most broadly cited of these open system models is that developed by Jolly (1965) and Seber (1965, 1973).

The Jolly-Seber model involves repeated sampling of the population at successive time intervals. During each capture period, unmarked individuals are "tagged" and released, along with other previously marked individuals. Within this context, $\hat{N}$ is calculated as in equation [1], but subscripted to reflect a sampling dimension, such that:

$$\hat{N}_i = \hat{M}_i\left(\frac{n_i}{m_i}\right) \qquad [2]$$

and:

$$\hat{M}_i = n_i\left(\frac{z_i}{r_i}\right) + m_i \qquad [3]$$

where z_i is the number of individuals that are marked and released before sample i and are *not* recaptured in sample i but which *are* recaptured in a subsequent sample and where r_i is the number of marked members released from sample i that are subsequently recaptured. Two additional parameters define death (or out-migration) and birth (or in-migration) rates:

$$\hat{\phi} = \frac{\hat{M}_{i+1}}{(\hat{M}_i - m_i - n_i)} \qquad [4]$$

and:

$$\hat{\beta} = \hat{N}_{i+1} - \hat{\phi}_i \hat{N}_i \qquad [5]$$

The assumptions underlying this open system model are essentially the same as those associated with the closed system model discussed earlier. A thorough treatment of this material is provided by Jolly (1965), and additional information on the estimators in equations [2] through [5] has been provided by Seber (1965). There appear to be no published accounts describing the application of this method to the problem under examination here, although it has been presented as a potentially useful technique (Woodward, Bonnett, and Brecht, 1985; Brecht and Wickens, 1992).[2]

In some ways the application of admission based capture-recapture models to the problem of drug use estimation has great intuitive appeal. But the enterprise presents a number of unique challenges.

The first of these is that it is difficult to form an operational definition of the population being estimated when capture-recapture techniques are applied in this manner. This is due to a violation of the mark indelibility assumption that is inherent in both the closed and open system capture-recapture models. It comes down to this: A lake trout that is captured, marked, and released, remains a lake trout until the day it dies. An addict, on the other hand, does not remain in a constant state of drug use following treatment and discharge. He or she may resume use immediately, may stop permanently, or may resume use following some period of abstention. Nor do individuals remain in a constant state of drug use *prior* to capture, for that

[2]The Jolly-Seber model was examined within the context described in the preceding note using a routine provided in Davies (1971).

matter. Drug use is, in fact, episodic, and technically members have no probability of entering treatment, or being recaptured, so long as they refrain from drug use. This problem cannot be regarded in terms of death, or out-migration, as described within the context of the open system model, since individuals who are treated, released, and temporarily abstain may be effectively reborn, or reenter the system, in a "ready-marked" state should they resume use. Thus, if admission based capture-recapture analyses are to be applied to this estimation problem, the phenomenon of drug-using behavior must be conceptualized in a manner consistent with the underlying assumptions of the model.

Another basic problem derives from the extent to which a legitimate analogy can be drawn between animal capture and release and treatment admission and discharge. There are a number of underlying assumptions associated with the capture-recapture model that must be considered at this point. One is that the duration of delay in the captive state is relatively brief: in fact, the model assumes that the process of capture, marking, and release is virtually instantaneous. Obviously some disjuncture occurs when we attempt to equate captivity with treatment, although admittedly a substantial number of individuals remain in care for comparatively brief periods of time. Another is that captivity is essentially a random event, but there is a fundamental distinction that must be made between help seeking behavior, which is volitional, and capture, which requires only passivity on the part of an animal under study. As we shall see, this distinction has profound consequences in terms of limiting the applicability of admission based capture-recapture methods to the task at hand.

For the purpose of exposition, we will examine a system of drug using behavior and treatment activity where drug using behavior is defined in a manner as consistent as possible with these assumptions. Within this context we will then explore specific problems associated with the use of admission based capture-recapture models.

In keeping with the capture-recapture metaphor, we present a "two pond" model of drug use and treatment activity in Figure 1. To address the first operationalization problem described above, we propose the use of an "epoch" construct. An epoch is assumed to begin with an individual's first use of a given drug and to end with the last. As such, it subsumes periods of heavy drug use, relatively infrequent use, and temporary abstinence. In essence, it establishes the temporal boundaries of a drug using career. All individuals in the "drug users" pond are defined as having active drug using career. Since the size of the treatment population is known, it is the size of the population in the "drug users" pond that is the object of study. This conceptual device provides some level of dimensional integrity in terms of allowing us to understand precisely what population it is that we will attempt to estimate. It obviates directly the problems associated with the episodic nature of drug use and with the fact that individuals may lose their marks simply because they have stopped using drugs but may reemerge as marked when they resume use.

We note at this point that an alternative conceptualization is possible where individuals in the "drug users" pond are defined as in a state of drug use. This system would treat cessation of drug use as death, if permanent, and out-migration, if temporary. This scheme would also entail allowing in-migration of marked members in cases where such individuals resume use following a period of temporary cessation. Since describing this system is somewhat more difficult and since the conclusions that can be drawn in terms of identifying the problems that arise when admission based capture-recapture models are applied using either operationalization are identical, we speak from the more simple epochal system presented in Figure 1.

Figure 1

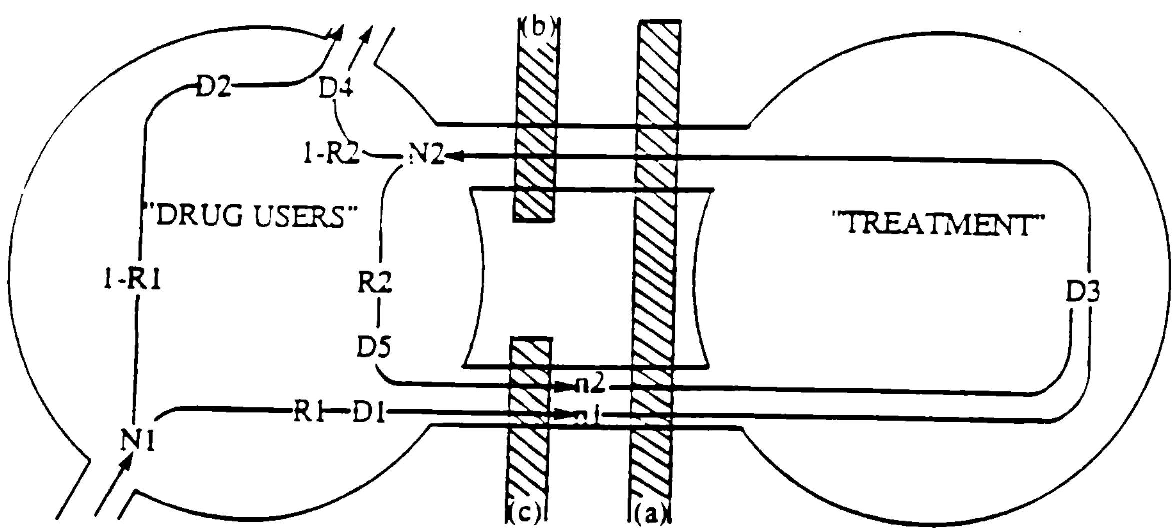

Figure 1 depicts a process wherein cohorts of individuals move through various states with known rates of transition and durations of delay. For simplicity, we will assume that the sizes of cohorts and all rates of movement and durations of delay are constant over time. Within this context, a group of individuals, N1, become initiates to drug use and enter the "drug users" pond. Of these, N1R1, a proportion bounded by 0 and 1, eventually seek treatment. These individuals experience a delay, subsuming periods of drug use as well as abstinence, which we define as D1, prior to entering treatment. Other individuals, N1(1-R1) within this context, refrain from seeking treatment for the entire duration of their epochs. These individuals remain in the "drug users" pond for a duration we define as D2. Individuals who eventually seek treatment following initiation of drug use remain in care for some period, depicted in Figure 1 as D3, and are then released as N2. Some of these, defined here as N2(1-R2), resume and eventually conclude their epochs without ever again seeking treatment. Such individuals remain in the "drug users" pond for a distribution of time defined here as D4, which is assumed to have a lower limit of 0. This acknowledges the possibility of a permanent cessation of drug use following treatment. Others, N2R2 in this case, eventually reenter treatment following a duration of D5.

If we imagine cohorts of individuals moving through the system described in Figure 1 in this manner, again making the simplifying assumption that all rates and delays remain constant over time and are not contingent upon the characteristics of prior states, we can envision the system "pumping up" so that population equilibria are eventually reached for the various groups defined in the "drug users" pond as well as in the "treatment" pond. We symbolically depict these equilibria for the remainder of this section simply as the number of individuals in an incidence (N1), or release (N2) cohort, times a corresponding rate and delay. Following this definition N1R1D1, for example, is the steady-state population level for the number of never-served individuals in the population in an epochal status who will eventually present for treatment. The question now becomes whether a net can be cast in some manner that allows us to mark and release members so as to infer the number of individuals in the "drug users" pond. A number of problems become immediately apparent.

The first of these is that when capture-recapture models are applied in this manner, nothing may be known about individuals who initiate and conclude drug use epochs without ever having sought treatment. In fact, using the notation developed above, N1, R1, and R2 *all* remain unknown. This is visually apparent in Figure 1, and the condition obtains regardless of how or where we place our net. Thus, admission based capture-recapture models do not make use of any information that would allow the size of N1(1-R1)D2 or N2(1-R2)D4 to be inferred. This observation highlights a basic and fundamental problem in the application of admission based capture-recapture models to the task of drug use estimation. When a sample of fish are captured, marked, and released at time 1, and another sample are captured and examined at time 2, then the ratio of marked to unmarked individuals in this second sample may be assumed to be representative of that which exists in the general population, subject to the condition of random dispersion. This is the essence of the methodology, which is lost when the entities under study must, deliberately and with premeditation, seek out the net.

While this observation suggests that admission based capture-recapture models are not appropriate for estimating the size of the total drug using population, the possibility remains that they may be of use in estimating the size of the treatment-bound population, more specifically, N1R1D1 and N2R2D5. This is feasible since, at equilibrium, the number of never-served individuals entering treatment during a given period may be taken as the per interval input to D1, and the number of previously served individuals is similarly the per interval input to D5. We denote these quantities henceforth as n1 and n2, respectively.

Within this more limited realm of possibility, another problem becomes apparent which seems generally to mitigate the applicability of admission-based open system capture-recapture models to the problem at hand. This derives specifically from delays that exist in the system. Open system models were developed in the biological sciences for studying population dynamics, including changes in the sizes of populations as well as changes in birth, death, and migration rates over time. With the introduction of D1 and D5, this capability disappears. This is due to the fact that variation in N1R1 and N2R2 can be observed only following delays of D1 and D5, respectively. Further, changes in the ratio of marked to unmarked members in the net may be due to variation in D1 and D5 per se, and not to variation in N1R1 and N2R2. This problem results directly from the distinction discussed above between volitional help seeking and captivity, which is unrelated to volition. The condition persists regardless of net location and the actual mechanism of marking and recapture that is employed. We thus discontinue the consideration of admission based open system capture-recapture models, as a class.

Limiting our focus now to the capability of an admission based closed system model in estimating the size of a treatment seeking population, we consider the issue of net location. Two methods of netting are presented in Figure 1. The first of these, depicted as (a), is a representation of the technique applied by both Frank et al. (1978) and Woodward, Bonnet, and Brecht (1985). It essentially involves treating all individuals admitted and released during a certain period of time, usually six months, as marked, and then examining the representation of such individuals in a second cohort, defined in an identical manner but admitted and released during a subsequent period of time, normally following a lapse of six months. This approach to net operationalization has a number of interesting consequences, not the least of which is the fact that given the delay in treatment (D3) and the delay for marked members who will eventually return (D5), only individuals with extremely short stays in each status will be represented in the net. To the extent that such individuals are more likely to seek treatment repeatedly than those who have longer stays, the ratio of marked to unmarked members will be inflated, and any population estimates will be negatively biased as a consequence.

Another basic problem here is that when capture and recapture are defined in this manner, a compromise must always be struck between the duration of time that separates release and recapture and the diminution of marked elements in the population that may be assumed to occur during this period due to the permanent cessation of drug use. The fact of the matter is that the system is not closed, and ultimately some assumptions about R2 and D5 must be made when estimates are derived in this manner. Frank et al. (1978) recognized this and made use of an arbitrary .33 "inactivation" factor in their analysis.

An alternative approach, which appears in the zoological literature but has not been applied to the drug use estimation problem to date, involves a procedure referred to as injection. An admission based application of this process is represented in Figure 1 by nets (b) and (c). Here, a cohort of individuals is marked and released at net (b), and collected over time in net (c). The technique, if applied in this manner, would eliminate the problems posed by D3, and could be used to provide direct estimates of R2 and D5. But it would require a relatively long follow-up period before the dynamics of the recovery process could be sufficiently understood so as to allow the scope of the population in the "drug users" pond to be estimated.

Examining now the feasibility of admission based injection marking and recovery for the purpose of estimating the size of a treatment seeking population, we offer a final observation which should probably be viewed, given the preceding discussion, as a coup de grace. As noted in the introduction to this section, a fundamental premise of all capture-recapture models is that the ratio of marked to unmarked elements in the net is representative of the numbers of marked and unmarked individuals in the population. As a practical matter, this can only be true if the distribution of delay between initiation of use and admission to treatment (D1) is identical to the distribution of delay between release and readmission (D5) for marked individuals who return to the net.

To demonstrate this point let us consider a simplified case where 10 individuals per month initiate use and eventually enter treatment following a delay of exactly 10 months and where five individuals are released each month and return following a duration also of exactly 10 months. The corresponding equilibria, or steady-state levels, for these populations would be 100 and 50, respectively, and the ratio of unmarked to marked members in the "drug users" pond would be 2:1. Individuals would be represented in net (c) in exactly this same proportion, which is determined solely by n1 and n2. If the delay for marked individuals had been five months rather than 10 months, however, then the corresponding population levels would be 100 and 25, a ratio of 4:1, and this would be at substantial variance with the 2:1 ratio that would ultimately appear in net (c). In fact, this differential between D1 and D5 is real and verifiable. Based on an analysis of data derived from admissions to methadone maintenance programs in New York State during calendar year 1989, Simeone, Nottingham, and Holland (1993) report findings indicating that the D1 distribution is lognormal with a mean of 103.3 months, and that the D5 distribution is negative exponential with a mean of 17.5 months.

Truncated Poisson Models

The Poisson distribution has a long and venerable history in the field of statistics. It has been used generally to describe a process wherein elements of a particular population may experience some discrete event, such as having an auto accident, buying a particular article of clothing, or as in this case, seeking treatment for drug addiction. The probability of such an event is usually denoted as α, and, in the simplest case, the model assumes that once an event occurs the probability of a second event remains α. The number of events experienced by an entity during a given observation interval is characterized in terms of state changes.

When viewed within the context of the problem at hand, applying the Poisson model involves examining treatment admission information for individuals over a given sample interval, say one year. During the year some individuals will have only one treatment episode, others will have two, and so forth. As noted above, transition among these states is governed by α and is assumed to be unidirectional.

When the interval of measurement is normalized as 1, as is usually the case, the form of the Poisson distribution, where p_i is the proportion of individuals in a given state, is simply:

$$p_i = \frac{\alpha^i e^{-\alpha}}{i!} \qquad [6]$$

In actuality, the number of people who use drugs during an observation interval and who do not present for treatment is unknown. But if an assumption is made that drug use and treatment follows a Poisson process, then admission information may be treated as limited, or, to be more specific, left truncated, and the size of the zero state can be estimated using a variety of techniques. This is the approach adopted by Sheridan, Levine, and Ney (1983), in which they developed a truncated form for the Poisson distribution--simply equation [6] divided by p_i for the zero state.

Like the admission based capture-recapture models reviewed in the preceding section, the admission based truncated Poisson model has a certain intuitive appeal. It is predicated on the simple assumption that observed admission activity may be regarded as a truncated sample of eventual admission activity for individuals at risk during the study interval. As such, the Poisson model is intended to provide a prevalence estimate relative to this period.

A number of problems arise, however, when we more closely examine the application of an admission based Poisson model to the task at hand. Since the techniques share the same mathematical underpinnings, it is not surprising that many of these are similar to the concerns raised in the preceding section on admission based capture-recapture analyses, and the "two pond" model presented there will continue to serve as a reference for our discussion.

The first issue is in some ways reminiscent of the problem of mark indelibility that we discussed earlier, although here we are concerned not with whether entities maintain their marks but more specifically with the extent to which they remain at risk throughout the duration of an epoch of drug use. As noted above, epochs are defined in a manner which includes all time spent in the "drug users" and the "treatment" ponds. Even within this limited domain they subsume periods of heavy use, relatively infrequent use, and temporary abstinence. The issue that must be examined within this context is whether it is appropriate to apply the Poisson model to a phenomenon with an underlying dynamic that is believed to effect changes in α over the course of time and, more specifically, during the course of an observation interval.

As a practical matter, α varies over the course of the observation interval for all applications of the Poisson process that we have examined. This condition obtains when we consider both physical and social phenomena; two references described by Coleman (1964) support this point. In the first well-known application of the Poisson process, Bortkewitsch examined the number of deaths due to horse kicks for ten corps in the Prussian Army over a period of 20 years, aggregating the data within one-year observation intervals. It is commonly believed that the assumptions underlying the Poisson process were not violated in this application, and yet it is clear that α was zero when the horses were left unattended each evening. A similar application involved bomb hits in small (0.25 square kilometer) areas of London during World War II (see Feller, 1957). The model was found to be applicable here as well, yet α was clearly zero when German planes were grounded by heavy weather.

Constancy of α is thus a matter of degree, or more accurately, a function of the level of resolution with which it is examined, just as the smoothest surface will exhibit crevices when viewed with an electron microscope. The task at hand in our case involves determining whether the probability of seeking treatment is relatively constant when viewed within the narrow context of an observation interval. Generally, we can safely assume that this probability is constant if the observation interval allows us to capture a representative sample of the periodicity of drug use. There is unfortunately little empirical data that can be brought to bear in this regard, but extant ethnographic studies suggest that the assumptions of the Poisson process are not necessarily violated by applying the model in this manner.

A second, and related, issue involves the presumed insensitivity of α to the actual occurrence of an event, in this case, treatment. As noted above, a basic assumption of the simple Poisson model is that every transition between states is governed by α, and that the process continues to be governed by α following the occurrence of an event. Within the present context, D1 and D5 may be interpreted as delay distributions associated with state changes resulting from a Poisson distributed arrival process.[3] Thus, the variability that is known to exist between D1 and D5 suggests that α is in fact sensitive to treatment. In fact, additional analyses conducted by Simeone, Nottingham, and Holland (1993) suggest that arrival may be governed by two distinct processes, one for first admissions and another for all subsequent admissions. In addition, findings reported by researchers working at the New York State Division of Substance Abuse Services suggest that following the first admission the rate of transition from one state to the next accelerates as a function of the number of prior admission events. This constitutes a basic violation of the simple Poisson process, but it is a phenomenon which can be accommodated through the use of what Coleman (1964) describes as a "contagious Poisson" model.

The contagious Poisson model is identical to the Polya distribution (see Feller, 1943) and is based upon the assumption that state changes affect subsequent probabilities of transition. The application of the contagious Poisson model would thus be appropriate if we believed that the probability of a drug user seeking treatment during an observation interval is a linear function of the number of preceding admission events occurring before the observation interval. In fact, this process seems to be consistent with the available information on the dynamics of admission behavior. It is a big step for an individual to acknowledge that he or she has a drug problem and needs professional care, but having come to this realization once, the process of seeking help becomes easier in the event that relapse occurs following discharge.

The contagious Poisson model incorporates a second parameter, β to reflect the variability in transition probability associated with prior state changes. The proportion of individuals in state p_i is given by:

$$p_i = \frac{\alpha(\alpha + \beta) \ldots (\alpha + [i - 1]\beta)e^{-\alpha t}(1 - e^{-\beta t})^i}{i!\,\beta_i} \qquad [7]$$

[3]Poisson arrival processes involving exponential delays are among the most studied queuing models in operations research (see Moder and Elmaghraby, 1978).

Its truncated form is simply equation [7] divided by p_i for the zero state.[4]

As presented above, an admission based model, appropriately constructed so as to include two distinct arrival processes, one simple and the other contagious, could conceivably provide a viable mechanism for estimating the size of an treatment seeking population during a given period of time. But to the extent that information is collected at admission (at net (c), referring again to Figure 1), a series of delays has been introduced into the admission-readmission process. Epochs are thus punctuated by episodes of treatment, and this has two consequences, one of which is fatal.

The first **effect of treatment episodes on our estimate** is that time in treatment is variable and zero-sum with respect to time at risk during the sample interval. As a practical matter it may be possible to accommodate this problem in the same manner as was described in the main text for dealing with population heterogeneity. This would involve a reformulation of the model to allow for variable time at risk during the observation interval or, equivalently, the derivation of estimates of α and β for members of each of several time-at-risk subpopulations.

The second effect of treatment episodes on our estimate is that the introduction of elements of D3, time in treatment, into the sample interval effectively serves to confound information on D1 and D5 with information on D3. This means that to legitimately infer the size of the zero state in a manner which excludes people in treatment, we must adopt a data coding convention in which an event only "counts" when both admission and release occur during the sample interval. Ultimately, this produces not only truncation of the zero state but also right censoring of individuals who are admitted during the year but never released. Because time in treatment is often substantially longer than the observation interval, this censoring would be extensive. This fact, in and of itself, precludes application of admission based truncated Poisson models to the task at hand.[5]

---

[4] Simeone, Nottingham, and Holland (1993) actually applied this model during the course of their research, and information on the derivation of maximum likelihood estimates for α and β as well as E-M based convergence routines are available from the authors.

[5]The preceding two sections have discussed issues in the application of admission based capture-recapture and truncated Poisson models to the drug use estimation problem. For the most part, these difficulties arise because we are attempting to apply existing techniques using only information collected at admission to, and release from, treatment. The question thus becomes whether the heuristic devise

presented in Figure 1 allows us to formulate an admission based model of any kind that is, from the outset, consistent with the process of drug use and help seeking behavior depicted there. The answer is a qualified yes, since the estimation of the parameters of such a model is likely to be problematic, given the nature of data that are currently available.

If we again consider the most elementary situation, one in which the system depicted in Figure 1 is at equilibrium, and thus where all rates of input and durations of delay are constant over time, then the size of the treatment susceptible population is by definition n1D1 + n2D5, using the notation developed above. More generally, however, and translating to conventional notation:

$$n1D1 + n2D5 = n1\sum_{i=1}^{n}(1-p1_{ci}) + n2\sum_{i=1}^{m}(1-p2_{ci}) \qquad [8]$$

where $p1_i$ and $p2_i$ correspond to discrete delay distributions, expressed as proportions of individuals in a given cohort who are admitted to treatment at each of n (or m) possible intervals, and where $p1_{ci}$ and $p2_{ci}$ are the respective cumulative proportions.

In the real world, of course, the situation is not this simple. All of the material presented thus far has focused upon estimating the parameters of a system at equilibrium. This approach, while useful in facilitating discussion, involves significant oversimplification. When the system depicted in Figure 1 is placed within a real historical context, we find that the existence of the "drug users" pond temporally precedes the existence of the "treatment" pond, at least so far as the "treatment" pond is operationally defined based upon the existence of a data base on client admissions. This means, reverting to our original notation, that each D1 distribution associated with an N1R1 incidence cohort is in some unknown way attenuated solely as a function of the creation of the "treatment" pond.

Even if this were not the case, it is possible that the duration of delay between first use and first admission would change in a cohort-specific manner which is not captured here. In fact, given the multiplicity of exogenous factors that are likely to affect help seeking behavior, there is no particular reason to believe that this is not the case. Information available on D5, but not presented here, suggests that variation exists in this distribution of delay as well. Joint variation in cohort size and corresponding delay distributions would have profound effects on estimates of N1R1D1 and N2R2D5. These observations argue against the use of equation [8] as a means of estimating the size of the treatment susceptible population. It remains, at a minimum, a useful mechanism for examining the size of the treatment bound population under varying sets of assumptions regarding plausible rates of admission and durations of delay.

A more realistic model, one that would be quite isomorphic with the process under study, would involve locating the system in time, and subscripting equation [8] in a manner which provides a unique D1 distribution for each N1R1 cohort and, similarly, a unique D5 distribution for each N2R2 cohort. This involves formulating two equations:

$$P1_t = P1_{t-1} + N1_tR1_t - n1_t \qquad [9]$$

and:

$$P2_t = P2_{t-1} + N2_tR2_t - n2_t \qquad [10]$$

where $P1_t$ and $P2_t$ are populations of treatment bound new admissions and readmissions, respectively, at the end of period t. Thus:

Summary and Conclusion

The preceding sections have examined the application of admission based capture-recapture and truncated Poisson models to the drug use estimation problem. Our conclusion is that when the collection of information on admission and discharge occurs as a result of the process of admission and discharge, it is not possible to develop unbiased estimates for the parameters of either the capture-recapture or truncated Poisson models. Thus we recommend that alternative methods of estimating these parameters be developed.

$$n1_t = \sum_{j=1}^{t-1} [N1_j R1_j (D1_j(t-j) - D1_j(t-j-1))] \qquad [11]$$

and:

$$n2_t = \sum_{j=1}^{t-1} [N2_j R2_j (D5_j(t-j) - D5_j(t-j-1))] \qquad [12]$$

where t is the month of admission, j indicates intervals relative to t, and all delays are cumulative distribution functions evaluated at t-j or t-j-1, as appropriate.

In fact, were we to calculate this and provide an additional equation for the size of the in-care population, here denoted as $P3_t$, such that:

$$P3_t = P3_{t-1} + n1_t n2_t - N2_t \qquad [13]$$

then we would have all the components of a basic model that could serve as a mechanism for demand estimation under various sets of assumptions. Unfortunately, equations [9] through [13] leave us with a mathematical formulation which corresponds reasonably well with the process under study but which requires information not likely to be available, even over relatively limited ranges of time. We should emphasize at this point that any estimates derived in this manner would be conditional to an unknown degree upon treatment capacity per se, and that this would be true of any estimation technique which made use of admission data in this manner.

REFERENCES

Bishop, Y.M.M., Feinberg, S.E., and Holland, P.W. 1975. *Discrete multivariate analysis: Theory and practice* Cambridge, MA: MIT Press.

Bonett, D.G., Woodward, J.A., and Bentler, P.M. 1986. A linear model for estimating the size of a closed population. *British Journal of Mathematical and Statistical Psychology* 39: 28-40.

Brecht, M.L., and Wickins, T.D. A comparison of multiple-recapture models for drug prevalence estimation.

Carothers, A.D. 1973. The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29:79-100.

Coleman, J.S. 1964. *Introduction to mathematical sociology*. New York: Free Press of Glencoe.

Cooley, P.C., Dunteman, G.H., Hamill, D.N., and Harwood, H. 1989. A review of methodologies for estimating drug use prevalence. Paper presented at What Works: An International Perspective on Drug Abuse Treatment and Prevention Research. New York City.

Davies, R.G. 1971. *Computer programming in quantitative biology*. London: Academic Press.

Feller, W. 1943. On a general class of contagious distributions. *Annals of Mathematical Statistics* 14: 389-400.

Feller, W. 1957. *An introduction to probability theory and its applications*. 2nd edition. New York: John Wiley and Sons.

Frank, B., Schmeidler, J., Johnson B., and Lipton, D.S. 1978. Seeking truth in heroin indicators: The case of New York City. *Drug and Alcohol Dependence* 3: 345-358.

Gilbert, R.O. 1973. Approximations of the bias in the Jolly-Seber capture-recapture model. *Biometrics* 29: 501-526.

Hser, Y., Anglin, D., Wickins, T., Brecht, M., and Homer, J. 1992. *Techniques for the estimation of illicit drug use prevalence: An overview of relevant issues*. Washington, D.C.: National Institute of Justice. Research Report No. NCJ 133786.

Jolly, G.M. 1965. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* 52: 225-247.

Moder, J.J., Elmaghraby, S.E. (eds.). 1978. *Handbook of operations research: Foundations and fundamentals*. New York: Van Nostrand Reinhold.

Phillips, D.T. 1971. Generation of random gamma variates from the two-parameter gamma. *AIIE Transactions* III, no. 3.

Rhodes, W. 1993. Synthetic estimation applied to the prevalence of drug use. *Journal of Drug Issues* (forthcoming).

Seber, G.A.F. 1965. A note on the multiple-recapture census. *Biometrika* 52: 249-259.

Seber, G.A.F. 1973. *The estimation of animal abundance and related parameters*. London: Griffin.

Shannon, R.E. 1975. *Systems simulation, the art and science*. Englewood Cliffs: Prentice-Hall.

Sheridan, J.R., Levine, G.L., and Ney, N.S. 1983. *1982 Maryland study of drug abuse incidence, prevalence, and treatment demand.* Silver Spring, Maryland: Maryland Drug Abuse Administration. Contract No. 52524.

Simeone, R.S., Nottingham, W.T., and Holland, L. 1993. Estimating the size of a heroin using population: An examination of the use of treatment admissions data. *International Journal of the Addictions* 78 (forthcoming).

Woodward, J.A., Bonett, D.G., and Brecht, M.L. 1985. Estimating the size of a heroin-abusing population using multiple-recapture census. Rockville, Maryland: National Institute on Drug Abuse. NIDA Research Monograph No. 57, 158-170.

APPENDIX II

SYNTHETIC ESTIMATION: SELECTED EXAMPLES

SYNTHETIC ESTIMATION: SELECTED EXAMPLES

In this appendix we summarize four recent applications of the synthetic estimation (SE) technique used to derive estimates of cocaine and heroin use. In doing so we illustrate that SE often relies on other techniques to obtain anchor points. Other reviews of SE applications are provided by Cooley and Liner (1990) and Spencer (1989). A selected bibliography is provided here that identifies both resource and reference material relevant to the development of SE models.

Rhodes (forthcoming)

In a study for the National Institute of Justice and the Office of National Drug Control Policy, Rhodes used cities that reported to the Drug Abuse Reporting System to estimate weekly cocaine and heroin use among users who are involved with the criminal justice system. Because many users are not involved with the criminal justice system, Rhodes complemented his estimates with other sources, including the National Household Survey on Drug Abuse and the High School Senior Survey. Thus, he derived anchor points for DUF cities.

He then used synthetic estimation techniques to derive estimates for non-DUF cities. For this purpose, he relied principally on arrests for drug-law violations and emergency room mentions for cocaine and heroin. He concluded that there were roughly 2 million weekly users of cocaine, and about one-third that number of weekly users of heroin, during the latter part of the decade.

Hamill and Cooley (1990)

The authors developed national estimates of the number of heroin users and of the number of heroin addicts based on an assumption that the number of heroin users is proportional to the number of emergency room (ER) mentions of heroin as a reason for seeking emergency medical treatment. ER mentions came from a probability sample of Drug Abuse Warning Network (DAWN) reporting facilities. The anchor points were modifications of prevalence estimates reported by Newmeyer (1988) for San Francisco and by Frank et al. (1978) for New York City. Hamill and Cooley estimated between 640,000 and 1,065,000 heroin addicts for

1987, the last year for which they provided estimates. If the ratio of ER mentions to heroin addicts held for 1990, during which there were 33,884 mentions of heroin, there must have been roughly 681,000 heroin addicts during 1990.

Wish (1990)

The author used synthetic estimation to place a lower limit on the number of weekly cocaine users. For data, Wish used the percentage of arrestees who tested positive for cocaine in the 20 cities that reported to the Drug Use Forecasting (DUF) program. Assuming that this percentage would apply in the other 41 large cities that had populations above 250,000, he multiplied the percentage of arrestees who tested positive in the DUF cities by the total number of arrestees in the 61 largest cities to derive an estimate of the total number of arrestees who would have tested positive for cocaine had all arrestees in the large cities been tested. This estimate was multiplied by 0.75 to account for multiple arrests (assuming an average of 1.33 arrests per arrestee); further, all those who tested positive were assumed to be weekly users of cocaine. Thus, Wish estimated 1.0 to 1.3 million weekly cocaine users were arrested in the 61 largest cities in the United States.

Although Wish's estimate is sensitive to his assumptions and to the accuracy of the DUF data, his estimate led many observers to the conclusion that weekly cocaine users must be far in excess of the 0.6 to 1.2 million estimated for the entire U.S. population based on the National Household Survey on Drug Abuse. Wish demonstrated that estimates do not have to be perfect to be useful.

The Senate Judiciary Committee

The Committee estimated the number of weekly cocaine users in 1988 (U.S. Senate, 1990a) and 1989 (U.S. Senate, 1990b). The Committee estimated weekly cocaine use among individuals entering treatment, among the homeless, among arrestees, and among the household population. The Committee's estimates were derived by adding the number of heavy users in these four groups and subtracting the estimated overlap.

Heavy users in treatment were estimated to be the number of people admitted to publicly funded treatment centers for whom cocaine was the primary or secondary drug of abuse. These estimates were adjusted to account for multiple treatment admissions for the same person during the reference year and to account for the prevalence of heavy use among this group (95 percent were assumed to be weekly users). No adjustments were made to account for admissions to private treatment facilities because data were lacking.

Drug use among the homeless was estimated by multiplying estimates of the number of people who were homeless by estimates of the percentage of homeless people who use cocaine weekly. The Committee reported that both estimates were highly judgmental but that few of the Nation's weekly cocaine users could be considered homeless regardless of the assumptions made about the homeless.

Extending Wish's analysis (discussed earlier), the Committee estimated the number of weekly cocaine users among the arrested population in large cities by multiplying the percentage testing positive for cocaine in the DUF sites by the number of arrests in the Nation's largest 61 cities. Then the Committee used conservative assumptions about the percentage of arrestees who would have tested positive at other locations--small cities, suburbs, and rural areas--to derive an estimate of the total number of arrestees who would have tested positive across the country. The Committee assumed that 80 percent (rather than Wish's 100 percent) of those who tested positive were weekly users. After making assumptions about the overlap among these three groups and between these groups and a residual core of heavy users who report to the NHSDA, the Committee concluded there were an estimated 2.2 million heavy users in 1988 and 2.4 million in 1989.

Summary and Conclusion

All of these SE approaches share a common method. First, all assume a linear relationship between an indicator variable and the number of drug users. In the studies by Hamill and Cooley and by Wish only a single relationship was used to derive estimates. (For Hamill and Cooley, the indicator variable was ER mentions for heroin; for Wish, it was arrests.) In the studies by the Senate Judiciary Committee multiple relationships were estimated and combined to produce estimates. Second, all studies used anchor points that were derived using a variety of techniques. The validity of these measures is subject to question. Third, none of the studies used probability theory to derive confidence intervals, but each study attempted to identify factors that influenced the accuracy of the reported results.

The proposed analysis plan described in the Data Requirements section of this proposal allows these characteristic limitations of SE methodology, as it has been applied in the area of drug use estimation, to be addressed. This is accomplished by developing accurate local estimates using the reformulated capture-recapture model and by effectively creating large numbers of anchor points based on relatively small geographic areas. This will allow the relationships between indicators and anchor points to be modeled with a substantially higher level of precision and the accuracy of the estimates to be improved to a corresponding degree.

Selected Bibliography

Blumstein, A., Cohen, J., Roth, J., and Visher C. (eds). 1986. *Criminal careers and "career criminals"*. Washington, D.C.: National Academy Press.

Bureau of Justice Statistics. 1989. *Sourcebook of criminal justice statistics 1988*. Washington D.C.: Department of Justice.

Bureau of Justice Statistics. 1990. *Prisoners in 1989*. Washington, D.C.: Department of Justice.

Chaiken, M. and Chaiken, J. 1985. *Who gets caught doing crime?* Paper prepared for the Bureau of Justice Statistics. Department of Justice. Washington, D.C.

Hamill, D. and Cooley, P. 1990. *National estimates of heroin prevalence 1980-1987: Results from analysis of DAWN emergency room data.* Research Triangle Institute.

Cooley, P. and Liner, E. 1990. *Heroin prevalence estimation, final report, task 1.* Submitted to the National Institute on Drug Abuse. Rockville, Maryland. Contract number 271-87-8302.

Elliott, D., Huizinga, D., and Menard, S. 1989. *Multiple problem youth: delinquency, substance use, and mental health problems.* New York: Springer-Verlag.

Feucht, T. Undated. An analysis of drug use among female arrestees in DC: the use of cocaine among prostitutes. Department of Sociology, Cleveland State University, Cleveland, Ohio. Unpublished paper.

Frank, B., Schmeidler, J., Johnson B., and Lipton, D.S. 1978. Seeking truth in heroin indicators: The case of New York City. *Drug and Alcohol Dependence* 3: 345-358.

Goldkamp, J., Gottfredson, M., and Weiland, D. *The utility of drug testing in the assessment of defendant risk at the pretrial release decision.* Department of Criminal Justice, Temple University. Unpublished draft.

Goldkamp, J., Jones, P., Gottfredson, M., and Weiland, D. *Volume II: Assessing the impact of drug related criminal cases on public safety: drug related recidivism.* The Project to Assess the Impact of Drug Related Criminal Cases on the Judicial Process, Jail Overcrowding and Public Safety; Temple University. Unpublished manuscript.

Goldkamp, J., Jones, P., and Gottfredson, M. Undated. Measuring the impact of drug testing at the pretrial release stage: Pretrial drug testing in Milwaukee County, New Castle County and Prince George's County. Preliminary report and evaluation update from the Pima County Program.

Gfroerer, J., and Brodsky, M. 1991. Estimation of drug abuse prevalence in California using the National Household Survey on Drug Abuse. Paper presented at meeting of the Sacramento Statistical Association, Sacramento, California, March 27.

Harrison, L. 1990. The validity of self-reported drug use among arrestees. Washington, D.C.: Department of Justice, National Institute of Justice.

Hser, Y., Anglin, D., Wickens, T., Brecht, M., and Homer, J. 1992. *Techniques for the estimation of illicit drug-use prevalence: An overview of relevant issues.* Washington, D.C.: National Institute of Justice. NIJ Research Report No. NCJ 133786.

Hubbard, R., Marsden, M., Rachel, J., Harwood, H., Cavanaugh, E., and Ginzburg, H. 1989. *Drug abuse treatment: A national study of effectiveness.* Chapel Hill: The University of North Carolina Press.

Hunt, D. and Rhodes, W. 1992. *The characteristics of heavy cocaine users.* Washington, D.C.: Office of National Drug Control Policy.

Hyatt, R. and Rhodes, W. 1992. Price and purity of cocaine: The relationship to emergency room visits and deaths, and to drug use among arrestees. Washington, D.C.: Office of National Drug Control Policy.

Johnston, L. and O'Malley, P. 1984. Issues of validity and population coverage in student surveys of drug use. In: Rouse, B., Kozel, N., and Richards, L. (eds.). *Self-report methods of estimating drug use: meeting challenges to validity.* Washington, D.C.: National Institute on Drug Abuse, NIDA Monograph 57.

Majchrzak, A. 1984. *Methods for policy research.* Newbury Park, CA: Sage Publications.

McCalla, M. and Collins, J. Patterns of drug use among male arrestees in three urban areas. Research Triangle Institute. Unpublished paper.

McNagny, S. and Parker, R. 1992. High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic. *Journal of the American Medical Association* 267, no. 8: 1106-1108.

Mieczkowski, T. 1992. The accuracy of self-reported drug use: an evaluation and analysis of new data. In: R. Weisheit (ed.). *Drugs, crime and the criminal justice system.* Anderson Publishing.

Newmeyer, J. 1988. The prevalence of drug use in San Francisco in 1987. *Journal of Psychoactive Drugs* 20, no. 2: 185-189.

Nurco, D. 1985. A discussion of validity. In: Rouse, B., Kozel, N. and Richards, L. (eds.). *Self-report methods of estimating drug use: Meeting challenges to validity.* Washington, D.C.: National Institute Drug Abuse, NIDA research monograph 57.

Rhodes. W. Using the drug use forecasting system to estimate the prevalence of heavy cocaine and opiate use. Report delivered to the National Institute of Justice. Washington. D.C. April 3, 1991.

Rhodes. W. and Hyatt. R. The price of illegal drugs. Interim report delivered to Office of National Drug Control Policy. May 15. 1992.

Rhodes. W. and McDonald. D. 1991. *What America's users spend on illegal drugs.* Washington. D.C.: Office of National Drug Control Policy. June.

Rhodes, W. and Scheiman. P. Forthcoming. *What America's users spend on illegal drugs, 1992.* Washington. D.C.: Office National Drug Control Policy Technical Paper.

Rouse. B., Kozel, N., and Richards. L. 1985. *Self-report methods of estimating drug use: meeting challenges to validity.* Washington. D.C.: NIDA research monograph 57.

Rossi. P. Wright. J., and Wright. S. 1978. The theory and practice of applied social science research. *Evaluation Quarterly* 2: 171-191.

Schmeidler. J. 1991. Projecting sample results to substate areas with known population characteristics. In: Wallack. S. (Chair). *Resource materials for state needs assessment studies.* Washington. D.C.: Office of Treatment Improvements.

Simeone. R., Frank, B., and Aryan. Z. 1991. Social indicator models. In: Wallack. S. (Chair). *Resource materials for state needs assessment studies.* Washington. D.C.: Office of Treatment Improvement.

Spencer. B. 1989. The accuracy of estimates of numbers of intravenous drug users. In: Turner. C., Miller. H., and Moses. L. (eds.). *AIDS, sexual behavior and intravenous drug use.* Washington. D.C.: National Academy Press.

Smith. D., Wish. E., and Jarjoura. G. 1989. Drug use and pretrial misconduct in New York City. *Journal of Quantitative Criminology* 5, no. 2: 101-125.

Toborg, M., Yazer. A., and Bellassai. J. 1988. *Assessment of pretrial urine testing in the District of Columbia: analysis of drug use among arrestees.* Washington. D.C.: National Institute of Justice. NIJ Research No. NCJ 116902. microfiche.

Toborg, M.. Bellassai. J., Yezer. A., and Trost, R. 1989. *Assessment of pretrial urine testing in the District of Columbia.* Washington. D.C.: National Institute of Justice.

U.S. Senate. Committee on the Judiciary. 1990a. Hard-core cocaine addicts: measuring--and fighting--the epidemic. A staff report prepared for the use of the Committee on the Judiciary.

U.S. Senate. Committee on the Judiciary. 1990b. Drug use in America: Is the epidemic really over? Summary of findings of a majority staff study prepared for the use of the Committee on the Judiciary.

Wish. E. 1990. U.S. drug policy in the 1990s: Insights from new data from arrestees. *The International Journal of the Addictions* 25. no. 3: 377-409.

Wish. E., Cuadrado. M., and Martorana. J. 1986. Estimates of drug use in intensive supervision probationers: results from a pilot study. *Federal Probation.*

Wish. E. and O'Neil. J. 1989. *DUF: Drug Use Forecasting, January to March 1989.* Washington. D.C.: National Institute of Justice. Research in Action.

Yazer. T. 1990. Characteristics of the juvenile urine-testing sample. Memo. March 29.

APPENDIX III

FIELD SURVEY TECHNIQUES FOR "STREET" POPULATIONS

FIELD SURVEY TECHNIQUES FOR "STREET" POPULATIONS

As was briefly discussed in the body of this paper, the most active users of illegal substances are not likely to appear in traditional survey sources for several reasons. They are often homeless or in transient living situations which by definition place them outside the definitional boundaries of the NHSDA. They are also often involved in crime and are incarcerated for some periods of time. Finally, they are involved in an illegal activity, drug use, and are not always forthcoming about their participation in this activity when surveyed.

Traditionally, many drug researchers have looked to drug treatment populations to describe hard-core drug users, and, in fact, a large number of hard-core users rotate in and out of treatment at some point in their drug use careers. The rate of this movement could be determined through a detailed examination of treatment records. However, as data from the recent NADR projects have shown, a surprising number of hard-core drug users never enter treatment. Almost 40 percent of this population (50,000 hard-core drug users sampled by the NADR project) had never been in treatment before and were not currently in treatment.

What is a "hard-core" user? A definition of hard-core drug use can be derived from a framework of absolute frequency of use, for example, any use beyond a certain frequency or amount is likely to be problematic and thus defined as "hard-core," or from one of dysfunction, for example, DSM-III-R criteria, self-admission of difficulty or treatment seeking behavior. We are choosing a frequency/amount of use definition for several reasons. First, administration of the DIS to a large sample of users in order to make the DSM-III-R diagnosis of dysfunction related to substance abuse is unrealistic. Second, admission of dysfunction either through treatment seeking behavior or self-admission is not a reliable measure of serious drug abuse. Many serious users do not admit difficulties until they have bottomed out or are ready for treatment even though difficulties are present, so that self-realization of difficulty is more a measure of treatment readiness that a description of status. Finally, the frequency with which someone uses a drug is an important policy variable. Put simply, high frequency users consume more drugs and represent the driving force behind the drug market. They are also more likely to be involved in crime to support larger drug purchases and to need treatment services at some point in time than a casual or infrequent user.

Therefore, the questions for this exercise are where can hard-core users be located, how do we find them, and what data can they provide that will help us develop the parameters for the complete model? In the sections which follow we will discuss some approaches to these questions.

Where are hard-core drug users located, and how can we find them?

The total population of heavy users will contain persons at varying stages in what is often termed a "career" in drug use. This career corresponds to both different ages of users as well as different stages of drug use: initiation, addiction, recovery, relapse, etc. For some users, the career is relatively brief; for others it lasts a lifetime. A sampling strategy needs to reflect persons at all points in their drug use career, with the possible exception of early stage use, or initiation. Unless the policy questions are confined only to drugs like heroin or cocaine, it also must be designed to sample from different types or populations of users and at different stages in their careers.

Institutional contacts are the easiest and least costly first sources. Treatment providers are an obvious source, one which was discussed earlier in the paper. However, heavy users may not access treatment. They do, however, use other public services through which they can be sampled. For example, heavy users often experience medical difficulties associated with drug use and, because of limited funds, use community health centers and emergency rooms for both emergency and nonemergency medical care. Developing a sample of drug users from urban emergency rooms is a logical first step and has been used successfully in other research (Shader et al., 1982).

In addition, community health centers can be important sources of sampling. In our research for NIDA designing and evaluating an demonstration project designed to reach pregnant substance abusers in Boston and Los Angeles indicated that a large number of women would not seek prenatal care through a regular hospital facility until delivery but did utilize neighborhood or community based care facilities. Similarly, clinics which specialize in health services related to drug use would merit special consideration for inclusion in the sample. These sites include clinics which provide testing and treatment for sexually transmitted diseases including testing for HIV infection, family planning services, and pediatric treatment.

Finally, programs which serve the homeless or indigent are also likely candidates for this sample: Women, Infants and Children (WIC) programs, shelters, public welfare agencies, programs for runaways, and crisis hotlines.

Because of the range of persons using these various institutions, the most commonly used illicit drugs (heroin, cocaine, sedatives, and hallucinogens) and all intermediate and late stages of use should be represented through this sampling strategy. For example, limiting sampling to emergency rooms might produce an over-representation of drugs which produce overdose or poisoning effects and/or users in late stages of heavy use who are experiencing difficulties. By adding in community health services as well as other public service programs, a better distribution of users and career stages will result.

Even canvassing these sources, we would miss some hard-core users. Users who did not avail themselves of public services (because they can afford private health care services or because they do not receive health care) would be under-represented. These are also the persons most likely to be under-represented in survey data sources. To include this group in the sample strategy, we will seek data from noninstitutional contacts.

The noninstitutional or "street-based" sample is unique. It is likely to be composed of the most acute drug users. Relying solely on this sample would bias the characterization of the total pool of active users in this direction, as it is likely to suffer from a narrow set of characteristics. However, used in conjunction with the institutional contacts suggested above, it should provide an exhaustive sample of active users.

Researchers in the field have long recognized the need to seek out hard-core drug users in noninstitutional settings, and drug users have been studied extensively using ethnographic field techniques (Preble and Casey, 1969; Hughes et al., 1971; Waldorf, 1977; Hanson et al., 1985 Adler, 1985). Since most users buy their own supply of drugs and distribution of that supply is generally concentrated in particular user "hangouts." This technique can include simple census interviewing of all persons in an area, snowball sampling networks of users (Biernacki and Waldorf, 1981), setting up storefront research operations, and recruiting users for research through outreach workers and even advertising.

These field techniques have been a mainstay of the drug research field for decades (see Hunt, 1989 or Atkins and Beschner, 1980 for a discussion of ethnographic techniques used to examine drug using and criminal behavior). The term "ethnography" refers to an analysis of behaviors using the definitions or the world-view of the persons under study. It uses methods of intensive field observation and interviews to gather data in the natural setting of those behaviors from a nonrandomly selected, or purposive or snowball, sample. Ethnography is particularly useful in preliminary research into new phenomena or when the appropriate measures are unknown and the individuals or groups to be studied and/or their behaviors are hidden. The techniques are not appropriate for drawing causal inferences but are most useful in examining patterns of behavior.

What data are needed for development of the model, and how can we obtain them?

Data collected in interviews in the field are self-report and heir to all of the problems inherent in measurement of criminal careers from self-report data. Since official records are also prey to a host of reliability and validity issues in this area, and since they are not available for those not accessing treatment, we are left with the task of increasing the accuracy of the measurements gained through self-report. This can be accomplished in several ways. First, it is critical to decrease the threat to the respondent through establishing rapport, through ensuring anonymous and confidential interviewing, and through employing only trusted workers indigenous to the culture under study to locate and interview subjects. Second, problems of recall can be diminished through the use of a "memory structure" such as the timeline or calendar interview format discussed below. Third, internal consistency of responses can be checked through the overlay of period responses (for example, comparing times at liberty with crime commission). In addition, we will ask respondents to submit to drug testing to test the reliability of their responses about recent drug use.

One way to improve the accuracy of self-report information on drug use and other illegal activities is by employing a central feature of many ethnographic studies--the life history or life events interview. This interview reviews with the user their drug use and often criminal activity starting with the beginning of their drug use careers.

The interview in a timeline format was pioneered in the work of Nurco and colleagues who used it to improve recall of extended time periods by narcotics users (Nurco, Shaefer and Hanlon, 1985). The work of McGlothlin and colleagues (McGlothlin, Anglin and Wilson, 1977) which used retrospective data collection techniques to recall events in extended time periods between interviews, has been expanded by Anglin and colleagues into the Drug Abuse Research

Center's Event History Interview, an intensive interview used to recall such things as crimes committed, drug use, periods of incarceration, and periods in treatment.

In these format, time units (months or multiple month periods) are displayed for each respondent on a grid. The interviewer, using standardized questions, or probes, locates major life events on the grid to orient the subject to specific time periods. He or she then asks the respondent to recall an activity (such as treatment) for each time unit. This technique has also been used successfully to increase recall accuracy in other interview settings (Freedman et al., 1988).

Using the timeline instrument, we are able to gather information from the sample of users not captured in treatment populations to determine their probability of "capture" into that population and hence an estimate of the total number of active drug users. The kinds of data needed in this effort--duration and intensity of periods of drug use (by drug), abstinence, and treatment--are familiar to experienced drug researchers and both commonly and successfully asked of hard-core users in field studies. Because of the informal setting (unlike the settings respondents to the NHSDA--who are interviewed in their homes--or respondents to DUF--who are interviewed in a compromised situation, in a lockup or booking facility) and the guarantee of confidentiality, data collected in this manner are likely to be more reflective of the actual behaviors of hard-to-reach hard-core drug users. The extremely heavy drug use reported throughout the NADR projects in such interview settings is testimony to the effectiveness of this technique. It is important to bear in mind that while some relatively generous margin of error is expected in interviews based on problems of recall or fear of disclosure or exaggeration, absolute precision in the dates provided is not necessary for the data to be useful. Since we are collecting career data on a number of types of hard-core drug users (crack users, heroin users, etc.) in this method, we are able to use those data in conjunction with treatment data to determine the size of subsets of the hard-core user population.

The data needed to complete the model are fairly straightforward. We need to understand the patterns of admission and readmission to treatment by types of drug users over their drug use careers. We are also interested in the overlap between other institutional populations (for example, drug users in prison) and the overlap between our samples. We can obtain this information through simply asking respondents in each sample about such issues as their utilization of services, incarceration, hospitalization, and living arrangements over a specified time period. Finally, since we may find entry rates different for different types of users, we would also ask questions about patterns of drug and alcohol use and basic demographic information.

REFERENCES

Adler 1985

Atkins, C. and Beschner, G. 1980. Ethnography: A research tool for policymakers. *Drug and Alcohol Fields*. Rockville, Maryland, National Institute on Drug Abuse.

Biernacki, P. and Waldorf, D. 1981. Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods and Research* 10: 131-163.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., and Young-DeMarco, L. 1988. The life history calendar: A technique for collecting retrospective data. *Sociological Methodology* 18: 37-68.

Goldstein, P. 1979. *Prostitution and Drugs*. Lexington, Massachusetts: Lexington Books.

Hanson, W., Beschner, G., Walters, J., and Jaffe, J. 1971. The social structure of a heroin copping community. *American Journal of Psychiatry* 125, no. 5: 551-558.

Hughes et al. 1971

Hunt 1989

McGlothlin, W., Anglin, D., and Wilson, B. 1977. The follow-up admission to the California Civil Addicts Program. *Advances in Alcohol and Drug Abuse* 4: 179-199.

Nurco, D., Shaefer, J., and Hanlon, T. 1985. The criminality of heroin addicts. *Journal of Nervous and Mental Disease* 173, no. 2: 94-102.

Preble, E. and Casey, J. 1969. Taking care of business: Heroin users life on the street. *International Journal of the Addictions* 4: 1-24.

Shader, R., Anglin, C., Greenblatt, D., Harmatz, J., and Santo, J. 1982. *An emergency room study of sedative-sypnotics overdoes: A study of issues and treatment*. Rockville, Maryland: National Institute on Drug Abuse.

Waldorf, D. 1977. *Doing coke: An ethnography of cocaine users and sellers*. Washington, D.C.: Drug Abuse Council Publications.

APPENDIX IV

MATHEMATICAL MODELS FOR ESTIMATING DRUG USE CAREERS

MATHEMATICAL MODELS FOR ESTIMATING DRUG USE CAREERS

Our sample is composed of people who admit to current drug use. We asserted in the text that this sample is sufficient to derive the parameters of the drug use career. In this appendix, we support that assertion.

As before we assume that the i^{th} user initiates drug use at time T_{i1}^{b} and ends it temporarily at time T_{i1}^{c}. He or she starts heavy drug use again at time T_{i2}^{b} and ends it temporarily at time T_{i2}^{c}, and so on. Treatment occurs at times T_{i1}^{c}, T_{i2}^{c}, etc. Hereafter, we drop the subscript i. We also define T^{o} as the time when the interview is conducted.

A mathematical model is necessary to estimate the parameters of interest. We start with a simple model. The first episode of drug use extends from the initiation of drug use to the first treatment episode. We assume this spell has an exponential distribution with parameter λ. We assume also that any subsequent periods of heavy drug use have a duration that follows an exponential distribution with parameter λ. One of our goals is to estimate λ.

We assume that the time from treatment to relapse has an exponential distribution with parameter γ and that the parameter γ is constant across the population. We are also interested in estimating γ.

We can estimate these parameters since the observable events have well-defined probability structures. First, the probability that a randomly selected user has never been in treatment is:

$$P[\textit{never treated, currently using}] = e^{-\lambda(T^{o}-t_1^{b})}$$

Second, the probability that a randomly selected user has been in treatment once and has relapsed is:

$$P[\textit{treated once, currently using}] = \left[1-e^{-\lambda(t_1^{c}-t_1^{b})}\right]\left[1-e^{-\gamma(t_2^{b}-t_1^{c})}\right]\cdot e^{-\lambda(T^{o}-t_2^{b})}$$

Of course, the probability that a randomly selected user has been in treatment once and has not relapsed by time T^{o} is:

$$P[\textit{treated once, not currently using}] = \left[1 - e^{-\lambda(t_1^e - t_1^b)}\right]\left[e^{-\gamma(t^o - t_1^e)}\right]$$

Such an occurrence is unobserved because the sample is limited to current users.

We could write similar expressions for users who have been in treatment twice, in treatment three times, and so on. The point is that we have defined a probability structure for observable outcomes. Using maximum likelihood methods, we can estimate the parameters of interest to us, namely λ and γ. Using these parameters we can estimate Q, as used in formula [6].

This simple model can also be elaborated:

- Adopting the exponential distribution is unnecessary. We can substitute other assumptions (such as assuming the time until the first treatment episode is distributed as lognormal).
- λ and γ do not have to be constant across spells; their values could depend on past realizations of the process.
- λ and γ can be assumed to vary with characteristics of subpopulations of drug users. For example, we might write $\lambda = e^{X\beta}$, where X is a row vector of exogenous variables (gender, when drug use career began, and so on) and β is a column vector of parameters.
- We could introduce unmeasured heterogeneity into the model, for example, by rewriting $\lambda = e^{X\beta+\epsilon}$, where ϵ has a known distribution.
- We could introduce other events into the model. For instance, periods of time in jail and prison are likely for heavy drug users.

We could introduce these elaborations into formula [1]. The essential point is that we can estimate all the parameters of interest from a sample of admitted heavy users.

REFERENCES

Allison. P. 1982. Discrete-Time Methods for the Analysis of Event Histories. Sociological Methodology 1982. S. Leinhardt. ed. (Jossey Bass. San Francisco. California).

Flinn. C. and Heckman. J. 1982. New Methods for Analyzing Individual Event Histories. Sociological Methodology 1982. S. Leinhardt. ed. (Jossey Bass. San Francisco. California).

Lancaster. T. 1990. The Econometric Analysis of Transition Data. (Cambridge. Cambridge University Press).

Trussell. J. and Richards, T. 1985. Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure. Sociological Methodology 1985. N. Tuma. ed. (Jossey Bass. San Francisco, California).