Drug Use, Prostitution and the Prevalence of AIDS: An Analysis Using Census Tracts

Author(s): Richard A. Berk


Published by: Taylor & Francis, Ltd.

Stable URL: http://www.jstor.org/stable/3812775

Accessed: 18/05/2011 18:28
Brief Reports

DRUG USE, PROSTITUTION AND THE PREVALENCE OF AIDS: An Analysis Using Census Tracts

RICHARD A. BERK, Ph.D.
University of California, Los Angeles

This paper reports on preliminary efforts to study at the census tract level the reciprocal relationships between AIDS and street crime, particularly prostitution and drug offenses and the prevalence of AIDS. The Los Angeles County Department of Public Health provided the number of reported AIDS cases by census tract. The data on reported crimes and arrests, organized by reporting districts, came primarily from the Los Angeles County Sheriff’s Department and the Los Angeles City Police Department. Although this strategy needs replication and extension, the present findings suggest that sexual orientation, drug use and prostitution are useful risk variables at the census track level.

KEY WORDS: AIDS, prostitution, drug use, census tract

I. Introduction

This paper reports on preliminary efforts to study at the census tract level the reciprocal relationships between AIDS and street crime. Epidemiologists who study AIDS have, by and large, neglected the census tract as a useful analytical unit. And, while IV drug use and perhaps prostitution are related to the spread of AIDS, one might wonder, in addition, whether the nature and extent of narcotics use, prostitution, and related crimes such as burglary would change as IV drug users and prostitutes become concerned about contracting AIDS, or actually develop AIDS symptoms.

The results presented below, however, focus primarily on drug use and prostitution as AIDS “risk factors.” For reasons described later, data are not yet available to explore the “feedback” impact of AIDS on drug use, prostitution, and other kinds of street crime. Nevertheless, some of the key substantive issues in the feedback process will be briefly discussed.

II. Substantive Issues

Given the ways in which AIDS may be transmitted, the course of the epidemic is now being increasingly shaped by intravenous drug use and,

Data for this work were generously provided by the L.A. County Dept. of Public Health, The Los Angeles Police Dept. and the L.A. County Sheriff’s Dept. Thanks go to Mang King Cheung for help in data preparation and to David Freedman and William Mason for suggestions in the data analysis. Funding was provided the National Institute of Justice and the Rockefeller Foundation. This paper is a revision of the project final report written for the two funders. Requests for reprints should be sent to Dr. Richard A. Berk, Department of Sociology, UCLA, Los Angeles, CA 90024.
perhaps, prostitution (Anderson et al., 1986; Decker, 1987; Ginzburg, 1987; Kleinman, 1986; Nahmas, 1989; Turner et al., 1989). However, there is also good reason to suspect that, while street crime affects AIDS, the opposite holds as well. For example, with the threat of AIDS, many prostitutes may seek other sources of income. Some may turn to legal pursuits, while others may trade prostitution for shoplifting and theft. At the same time, many of their customers may begin to think twice about the risks involved in purchasing the services prostitutes provide.

Likewise, intravenous drug users may switch to drugs that need not be injected or may more readily seek out drug treatment programs (Des Jarlais & Friedman, 1988; Turner et al., 1989, pp. 202-209). Drug users experiencing AIDS symptoms also will be less capable of committing crimes to support their habits, and excess mortality may simply reduce the pool of prospective robbers and burglars. In short, there is some justification for predicting that property crimes linked to intravenous drug use will decline.

But the crime rate is affected not just by the activities of experienced offenders, but by the number of new recruits (Blumstein et al., 1986). Insofar as crimes are economically motivated, AIDS may increase incentives among some to steal and others perhaps to turn to prostitution. What might happen to a household, for example, if the sole wage earner is incapacitated by AIDS? Income is lost and medical expenses dramatically increased. Might not other household members be strongly motivated to find new sources of income, some of which may be illegal?

What about violent crime? There is ample evidence that a large portion of violent crime is committed by young men from less affluent urban neighborhoods (e.g., Tillman, 1986). A lot depends, therefore, on how many of them will contract AIDS. In certain locales, such as the New York metropolitan area, the number will probably be large. In the New York area, infection rates among intravenous drug users may well be over 50 percent (Newman, 1987; Turner et al., 1989, pp. 234-240), and most intravenous drug users are young men. The risks are compounded insofar as these young men have sexual contact with women who are themselves intravenous drug users. Finally, sexual transmission between young people who do not use intravenous drugs is no less a problem among inner city teenagers than among teenagers elsewhere.

In short, resentful young men, already with little to lose, may see their life’s chances further diminished. The motives for violence may increase, and the foundations for deterrence may be further undermined. Yet, AIDS symptoms and excess mortality may reduce substantially the likelihood that violent inclinations will be translated into violent acts. Trends in violent crimes will depend in part on the balance over time between these two processes.

To summarize, there is ample reason to suspect that AIDS and certain kinds of street crime are reciprocally related. However, there are both positive and negative feedback loops so that simple associations do not lead to simple

1Kane (1990) argues, however, that female prostitutes commonly have their customers use condoms, so that the major heterosexual risk for prostitutes is with their relationship partners.

2They may also practice safer drug use, such as sterilizing their "works."
dynamics. The ways in which AIDS and street crime are related, therefore, becomes an empirical question.

III. Research Methods

Research Design

The analyses to follow rest on a pooled cross-sectional and time series (or panel) design (Hsiao, 1986). The unit of analysis is the census tract, allowing for variation across tracts (i.e., cross-sections) and years (i.e., time). Both dimensions will be exploited in the analyses to be presented shortly.

For jurisdictions within the state of California, two kinds of crime measures are generally readily available: crimes reported and arrests. Such data may be obtained, organized by reporting district and time, from the California Bureau of Criminal Statistics (BCS) and/or particular police departments. The number of AIDS cases by census tract and time are available from many of the state's county health departments.

Given resource constraints, the research was initiated in Los Angeles County. Los Angeles County has obvious relevance; there is a large number of census tracts coupled with significant variation in crime and AIDS across tracts. Moreover, there are substantial populations of prostitutes and intravenous drug users. And the sample can be very large; yearly data can be collected for most of the 1980's for all of Los Angeles County's 1,634 census tracts.

Using census tracts as the unit of analysis has many advantages, not the least of which is the wide variety of data readily available. There is probably nothing even roughly competitive at the individual level. Equally important, however, there is a rich tradition in criminology of studying crime "rates" for particular geographical areas (e.g., Cohen & Felson, 1979; Cohen et al., 1980; Bursik & Webb, 1982). Sometimes such work reflects the simple aggregation of individual-level events, but it may also represent an effort to theorize about geographical areas per se. In either case, there are important ideas on which to build. As an empirical matter, moreover, criminologists typically do a far better job identifying regularities in crime data for geographical areas than in crime data for individuals.

Data Collection

The project received excellent support from the relevant public agencies in Los Angeles. The Los Angeles County Department of Public Health provided the number of reported AIDS cases by census tract and year in machine readable form. A clinical diagnosis was required for each reported case; a positive

\(^3\)While the data could be collected and organized by quarter, or even by month, the logistical burdens would be enormous. As will be described shortly, it is no small matter to combine AIDS data with crime data at the census tract level.

\(^4\)Reporting districts are typically smaller than census tracts, but may be aggregated into reasonable approximations of census tracts.
HIV test without a clinical diagnosis was insufficient. Likewise, a positive HIV test and an ARC diagnosis were insufficient.

Cases were located in census tracts by the residential address of the person with AIDS, with time defined as the date of diagnosis. It cannot be overemphasized that raw numbers underestimate somewhat the number of AIDS cases in a census tract and, of course, grossly underestimate the number of individuals who are HIV positive. And there is no doubt some lag between the onset of AIDS and an official diagnosis. In short, the AIDS data are subject to all of the problems complicating existing epidemiological research on AIDS (Abramson, 1988; Abramson & Rothschild, 1988; Isham, 1988). For present purposes, however, some of the difficulties are eliminated by, as a substantive matter, focusing on a diagnosis of AIDS rather than a positive HIV test. Moreover, the interest in relative variation across time and space means that reasonably stable and consistent measurement error may not cause insurmountable problems.

Unfortunately, in an effort to guarantee anonymity, the Los Angeles County Department of Public Health would not provide exact AIDS figures for census tracts with less than 5 cases. Observations for census tracts with less than 5 cases were reported as "0," meaning 4 or less. As the number of AIDS cases increases dramatically over the next 24 months, this problem will probably become unimportant. For the present, however, some distortions no doubt result. More will be said about this shortly.

The crime data present a somewhat different mix of opportunities and problems. On the one hand, it was possible to consider both reported crimes and arrests, organized into a large number of different crime categories. And, these data were readily available by year for the period of interest. On the other hand, the data were organized by reporting districts, not census tracts. Moreover, the data from Los Angeles County came from several different police agencies (although primarily the Los Angeles County Sheriff's Department and the Los Angeles City Police Department). Reporting conventions differ somewhat across agencies. Finally, the data were not available in machine readable form; the crime data had to be entered by hand.

Many months were spent constructing the required data files. In principle, the police reporting districts for both the county data and city data are easily aggregated into census tracts, since the boundaries of many reporting districts have been determined with census tracts in mind. The only significant problems resulted from changes over time in the boundaries of a few reporting districts and census tracts. By and large, however, these changes represented the division of larger census tracts and/or reporting districts into smaller census tracts and/or reporting districts (in response to population

---

5The more general problems with crime data are widely studied. For recent examples see Gove et al. (1985) and Cohen and Land (1984).
growth). It was relatively simple, therefore, to collapse newer areas into older ones.\textsuperscript{6}

Finally, the usual array of census data organized by census tract was available in machine readable form. Some of the data were available for 1980 and some were available for 1985.

\textit{Data Analysis}

It is important to emphasize again that the AIDS-crime relationship may well be reciprocal. Therefore, any examination of contemporaneous relationships between the prevalence of AIDS and street crime risks confounding two different causal effects, one from AIDS to crime and one from crime to AIDS. As a practical matter, therefore, one would ideally like to examine the lagged effects of AIDS on crime and the lagged effects of crime on AIDS. Since the latter are unlikely to appear until the number of people with AIDS increases substantially, the focus here will be on how street crime may be related to the prevalence of AIDS.\textsuperscript{7}

From graphs reported in earlier work (Berk, 1988) showing for Los Angeles County census tracts the number of AIDS cases by year from 1982 to 1987, it is clear that the epidemic began in the Hollywood area in 1983 and is still concentrated there. However, by 1985 and 1986, some AIDS cases could also be found in Santa Monica/Venice, Long Beach and south-central Los Angeles. In short, the spatial distribution of AIDS cases over time is clearly dominated by the presence of large concentrations of gays in certain areas of Los Angeles.

However, there are complications that are becoming more important as time passes. Parts of Hollywood are also well known for narcotics use and prostitution (male and female). In the two beach areas, there has been a substantial drug trade for some time. And Long Beach has had both prostitution and drug problems. In south-central Los Angeles, the epidemic is probably driven primarily (at least initially) by IV drug use and perhaps prostitution (but see Kane, 1990).

Whether with respect to street crime any of these considerations have empirical promise will probably depend on whether there are substantial relationships between street crime and AIDS. With that said, it is not at all apparent how best to proceed in the short run. First, even by 1988, most census tracts in Los Angeles County had no reported cases of AIDS, although some unknown number had between one and four cases. The analysis will be initially limited, therefore, to tracts patrolled by the Los Angeles Police

\textsuperscript{6}It is important not to minimize the amount of effort involved. All of the crime data had to be entered by hand and checked. With over 6,000 reporting districts, over 20 kinds of crime and 6 years, a lot of labor-intensive work was required. In addition, while it is simple in principle to write the programs required to construct the data set, a lot of code had to be written.

\textsuperscript{7}The use of simultaneous equation models, alternatively, requires better data and better theory than I have.
Department, where the vast majority of high AIDS tracts can be found. And, moreover, the correlation across census tracts between the number of reported AIDS cases and the actual number of AIDS cases is likely to be high. This would be fully consistent with the observations of public health authorities and the dominant transmission mechanisms in Los Angeles.

Second, there are also problems on the right-hand side of any explanatory equation. Arrests and reported offenses are imperfect measures of "real" crime, and the amount of crime is an imperfect measure of individual high-risk behavior. In addition, the number of AIDS cases in a census tract will be affected by far more than the amount of street crime; indeed, at this stage of the epidemic in Los Angeles, the role of street crime may well be unimportant, or at least unobservable.

Given these difficulties, it is perhaps best to recast the analysis a bit. Rather than thinking in terms of cause and effect, one can think of "explanatory" variables at the census tract level as proxies for risk. That is, the explanatory variables perhaps are at least associated with the true causal variables. For example, the number of reported drug-related crimes is probably associated with either the amount of IV drug use in a census tract or lifestyle patterns that may be associated with high-risk sex. Perhaps less directly, tracts with lower rents are probably those with more affordable housing, which in turn, may attract unattached young adults. Unattached young adults may be especially sexually active and less concerned about risky sex.

However, these sorts of relationships are vulnerable to the ecological fallacy: relationships at the census tract level do not necessarily hold at the individual level. Thus, even if (other things being equal) census tracts with more drug offenses have more AIDS cases, it does not necessarily follow that it is the individual drug users who are infected.

Finally, even by the summer of 1990, the "exposure source" in Los Angeles for the majority of reported cases of AIDS was male homosexual or bisexual behavior. The Los Angeles County Department of Public Health reported a total of 612 new cases through June of 1990, of which 60 percent were associated with homosexual or bisexual sexual contacts. For another 10 percent, the "exposure source" was a combination of homosexual or bisexual behavior and/or IV drug use. A little less than 10 percent were associated with IV drug use alone. Consequently, explaining the number of new AIDS cases in a census tract is substantially an exercise designating which census tracts have a large number of gay residents. Risk variables not linked to sexual orientation may not be productive until the middle of the 1990s.

8Complete datasets have been constructed for areas patrolled by the Los Angeles Police Department and the Los Angeles County Sheriff's Department. The data could not be easily pooled because the crime definitions were not comparable. And a separate analysis of the Sheriff's Department census tracts was seriously constrained by the very small number of census tracts with more than 4 AIDS cases (7 out of 424 tracts). More will be said about these issues shortly.
IV. Findings

The most recent and complete data readily available on the number of AIDS cases in particular census tracts is for 1988, and I will focus here on the number of new cases for that year. Of the 720 tracts patrolled by the Los Angeles Police Department (LAPD), 46 tracts had at least five new cases of AIDS reported. As Figure 1 shows, however, the distribution of AIDS cases (among tracts with at least five AIDS cases) is highly skewed. Thirty tracts have five to seven new cases but four have 15 or more. This implies that any analysis of variables associated with the number of reported cases of AIDS will be driven by a relatively small fraction of the census tracts. Put another way, the analysis may be primarily a reflection of binary difference between census tracts that have no new cases of AIDS and census tracts that have 5 or more new cases.

![Figure 1](https://example.com/figure1.png)

*Number of AIDS cases in 1988: For Tracts with cases.*

The two census tract risk variables of key interest are the reported number of prostitution offenses and the reported number of drug offenses. The number of prostitution offenses is most directly a proxy for high-risk sex and also may be related to IV drug use (since prostitution and IV drug use are often correlated). The number of drug offenses is most directly a proxy for IV drug use but also may be related to high-risk sex (since some drug users trade or sell sex for drugs, which are not necessarily taken intravenously).
Figure 2 shows a boxplot for the number of reported prostitution offenses for 1984. The year 1984 is used in part to allow for the lengthy AIDS incubation period. The geographical patterns are much the same for other years in the middle 1980s, however, so the same results follow from the crime data for any year in the early to middle 1980s.

The main lesson from Figure 2 is that the distribution for the number of reported prostitution offenses is skewed much like the AIDS data. This implies that any analysis of the relationship between prostitution offenses and the number of AIDS cases will likely have several very high leverage observations on which much of the substantive story will depend.

Moreover, it is important to keep in mind that prostitution offenses are rarely reported by citizens but derive primarily from initiative taken by police officers. As a result, the number of prostitution offenses is especially vulnerable to reporting biases introduced by variation in police practices. In addition, prostitutes and their customers do not necessarily live in the census tracts in which they meet; there is often a difference between where an arrest is made and where the offender lives (and where his or her case of AIDS is reported). In short, there is good reason a priori to be cautious about the potential explanatory power of the number of prostitution offenses.
Figure 3 shows a boxplot for the number of reported drug offenses in 1984. Drug offenses are generated by a combination of citizen reports and police initiatives, and drug users are perhaps more likely than prostitutes and their customers to live in the area in which their crimes are committed. The number of drug offenses, therefore, may be a more effective risk variable than the number of prostitution offenses. Again, however, the skewing is substantial, and several very high leverage observations are likely.

A high-risk census tract is likely to be associated with more than high offense rates for prostitution and drug use. Risk is likely to be higher in census tracts with high rates of sexually active people, especially if they have many partners and engage in particular sexual activities that increase the probability of HIV transmission. In short, lifestyle and life cycle matter.

In an effort to control for these factors, included as control variables are: 1) a quadratic function of the median age of males in a census tract, 2) median rent for residential housing in a census tract, 3) the arithmetic difference between the number of males and the number of females in a census tract, and 4) the natural logarithm of the total population in a census tract.

One might anticipate that the number of reported AIDS cases would be 1) higher in census tracts with higher concentrations of people in their 20's, 30's and 40's; 2) higher in census tracts with lower median rents (attracting young adults); 3) higher in tracts with a large number of men relative to women.

Figure 3
Boxplot: Drug Offenses.
(tracts with greater concentrations of gays); and 4) higher in tracts with more people at risk.\(^9\)

Figure 4 shows all of the scatterplots for each possible pair of variables: AIDS (number of AIDS cases), PROSTIT (number of reported prostitution offenses), DRUGS (the number of drug offenses), AGE and AGESQ (male median age and male median age squared), RENT (median rent), MOREMEN (the number of males minus the number of females), and POPUL (log of population). These are provided because of the extremely skewed nature of several variables; it is clear that whatever relationships exist among the key variables of interest depend on 20 to 30 census tracts.\(^{10}\) However, focusing on the first column (scatterplots with AIDS as one variable), it is difficult to see any compelling relationships, with the exception of the role of population.

\(^9\)The natural logarithm of population is used because it allows the dependent variables to be interpreted as a rate (i.e., AIDS cases per capita) when Poisson regression techniques are used (more on that shortly). However, this is not an essential point because one gets almost the same substantive story using population not logged.

\(^{10}\)It is also important to keep in mind that each scatter plot includes over 700 points, most of which fall on top of one another, often at the "low-low" corner.
Table 1 shows Poisson regression results.\textsuperscript{11} Since the right hand side is exponentiated, the regression coefficients are not easily interpreted (i.e., the dependent variable is in the metric of the log of the number of AIDS Cases). In the far right-hand column, therefore, the multipliers are provided in useful units. For example, the coefficient for males minus females is .0003, which translates into a multiplier of 1.003 for each additional person, or 1.03 times more AIDS cases per 100 people (1.0003 raised to the 100 power). That is, for each additional 100 males compared to females, the number of AIDS cases increases by a multiplicative factor of 1.03 times (or by 3 percent). Since the arithmetic difference between the number of males and number of females is a proxy for the concentration of gays in a census tract, the positive effect is fully consistent with expectations.

Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t Value</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-7.1000</td>
<td>-6.96</td>
<td></td>
</tr>
<tr>
<td>Prostitution Offenses</td>
<td>-0.0020</td>
<td>-3.16</td>
<td>0.82 per 100 offenses</td>
</tr>
<tr>
<td>Drug Offenses</td>
<td>0.0065</td>
<td>10.53</td>
<td>1.92 per 100 offenses</td>
</tr>
<tr>
<td>Median Male Age</td>
<td>0.2972</td>
<td>8.81</td>
<td>1.35 per year</td>
</tr>
<tr>
<td>Median Male Age Squared</td>
<td>-0.0026</td>
<td>-6.71</td>
<td>0.99 per year squared</td>
</tr>
<tr>
<td>Median Rent</td>
<td>-0.0055</td>
<td>-8.49</td>
<td>0.58 per $100</td>
</tr>
<tr>
<td># Males—# Females</td>
<td>0.0003</td>
<td>3.63</td>
<td>1.03 per 100 people</td>
</tr>
<tr>
<td>Natural Log of Population</td>
<td>0.2081</td>
<td>2.07</td>
<td>1.60 per log 100 people</td>
</tr>
</tbody>
</table>

The natural logarithm of population also operates as expected: more people, more cases of AIDS. Similarly, the anticipated results surface for age: AIDS cases increase with the median age of males in a census tract, but at a decreasing rate, and eventually begin to decline. The only surprise among the control variables is the effect of median rent. For each $100 increase in median rent, the Poisson regression coefficient is 1.03, which translates into a multiplier of 1.03 for each additional $100 of median rent.

\textsuperscript{11}Since the outcome variable is a count, one needs a non-linear regression function and disturbance distribution that cannot be negative. Poisson regression assumes that the log of the count is a linear function of a set of explanatory variables and that the disturbance distribution is conditionally Poisson. However, it also assumes that the events that make up the counts and the counts themselves are conditionally independent. All of these assumptions can be questioned for these data. However, the major problem is potential omitted variables; potential omitted variables are the weakest links in the analysis chain. In short, it would be at least inefficient to invest a great deal of time responding to such problems as spatial autocorrelations until more compelling and complete models are formulated.

The same arguments hold for the censoring problems in the data. In addition, while it is not especially difficult to write out what the likelihood function would be that would take into account the censoring (tracts with between 1 and 4 cases coded as 0), it is not clear what the properties of that function would be and, in particular, whether attempting to compute maximum likelihood estimates would lead to convergence at the global maximum. In any case, by ignoring the censoring, the reported regression coefficients are almost certainly biased conservatively toward zero (toward finding no effects).
rent, the number of AIDS cases is multiplied by 0.58. Other things equal, higher rent tracts have fewer cases of AIDS. Perhaps this should have been anticipated. In Los Angeles, tracts with relatively high concentrations of gays are also likely to be relatively high on other measures of affluence.

Turning to the crime measures, the number of drug offenses operates as expected. For each 100 offenses, the number of AIDS cases in a tract is multiplied by nearly 2.0. In contrast, the number of prostitution offenses has a fractional multiplier of 0.82. As noted above, however, there are a number of special problems with the prostitution variable.

To summarize, high-risk census tracts are apparently those with larger concentrations of gays, a larger number of sexually active individuals and a larger number of IV drug users. These are, of course, not mutually exclusive categories although they conflate the impact of the epidemic among gays and the epidemic among IV drug users. They also conflate, therefore, a higher SES epidemic with a lower SES epidemic. This may help to explain why tracts with a larger number of prostitution offenses seem to be at lower risk. Basically, two different kinds of processes are perhaps being combined.

As noted earlier, moreover, the analysis in Table 1 is essentially driven by the difference between tracts with no reported cases of AIDS and tracts with more than 4 cases of AIDS. One might wonder, therefore, which risk factors might surface if only tracts with five or more AIDS cases were examined. That is, within tracts clearly at risk, what makes for especially high risk? Table 2 shows the results.

Table 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t Value</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>7.796</td>
<td>0.86</td>
<td>—</td>
</tr>
<tr>
<td>Prostitution Offenses</td>
<td>0.0014</td>
<td>1.94</td>
<td>1.15 per 100 offenses</td>
</tr>
<tr>
<td>Drug Offenses</td>
<td>0.0004</td>
<td>0.59</td>
<td>1.04 per 100 offenses</td>
</tr>
<tr>
<td>Median Male Age</td>
<td>0.0539</td>
<td>1.75</td>
<td>1.05 per year</td>
</tr>
<tr>
<td>Median Male Age Squared</td>
<td>-0.0005</td>
<td>-1.86</td>
<td>0.99 per year squared</td>
</tr>
<tr>
<td>Median Rent</td>
<td>0.0011</td>
<td>1.02</td>
<td>1.11 per $100</td>
</tr>
<tr>
<td># Males—# Females</td>
<td>0.0000+</td>
<td>0.07</td>
<td>1.00 per 100 people</td>
</tr>
<tr>
<td>Natural Log of Population</td>
<td>-0.0266</td>
<td>0.30</td>
<td>0.88 per log 100 people</td>
</tr>
</tbody>
</table>

Given the small sample size (N = 46), any t values larger than 1.64 in absolute value (roughly the 0.10 level) are taken as worth considering, and two effects appear. Once again, the age distribution matters; the risk is greater in tracts with large concentrations of people in their sexually active years. In addition, the number of prostitution offenses now weighs in as expected: more offenses, more reported cases of AIDS. Yet, the mechanism remains obscure.

12 This analysis was suggested by Bill Mason.
Descriptively, such tracts are located in the Hollywood area that includes Santa Monica Boulevard, Hollywood Boulevard and the "Sunset Strip," where a relatively large number of gays reside and where all sorts of street life are highly visible. One might suspect, therefore, that the number of prostitution offenses in these tracts is really a proxy for other risk variables such as the number of young street people and a lifestyle rich in sexual activity and drug use (but not necessarily IV drug use).

In any case, however credible some of the findings may be, they are vulnerable to all of the usual concerns about omitted variables and overfitting. Ideally, the analysis would have benefitted enormously from a validation dataset.\textsuperscript{13} It would then have been possible to move the model in Table 1 to the new data and determine how well the model was able to identify tracts in the new dataset with a large number of reported AIDS cases (and those with a small number of reported AIDS cases). Had the model been able to properly identify the high- and low-risk tracts in the new data, the credibility of the findings would have been dramatically enhanced.

Unfortunately, the model's predictive ability, even within the "training" dataset, is not strong. While the censored and highly skewed nature of the outcome variable makes it difficult to simply summarize the model's fit, it is clear that most of the variability is not accounted for by the explanatory variables and that there is a general tendency to underestimate the number of AIDS cases in high-risk census tracts. And if the model does not "forecast" especially well within the training dataset, it will not do especially well in a new dataset; in fact, it will almost certainly do somewhat worse.

Table 3

\textit{Poisson Regression Results for Tracts Patrolled by County Sheriff's Department}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>$t$ Value</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.6742</td>
<td>-13.42</td>
<td>-</td>
</tr>
<tr>
<td>Sex Misdemeanors</td>
<td>0.0062</td>
<td>19.15</td>
<td>1.55 per 100 offenses</td>
</tr>
<tr>
<td>Drug Offenses</td>
<td>0.0029</td>
<td>3.54</td>
<td>1.34 per 100 offenses</td>
</tr>
<tr>
<td># Males—# Females</td>
<td>0.0016</td>
<td>10.75</td>
<td>1.17 per 100 people</td>
</tr>
</tbody>
</table>

As a fallback position, an attempt was made to reproduce the signs of the explanatory variables reported in Table 1. To this end, Table 3 shows Poisson regression results for Los Angeles census tracts patrolled by the Los Angeles County Sheriff's Department. Of the 424 tracts, however, only 7 had 5 or more reported cases of AIDS. It was impossible, therefore, to include more than a very few explanatory variables. Moreover, the crime categories in the Sheriff's Department data were a bit different from the crime categories used by the LAPD. In particular, "sex misdemeanors" is about as close as one can get to "prostitution offenses" (since most prostitution offenses are sex misdemeanors and vice versa).

Nevertheless, the results are encouraging. Drug offenses and males minus

\textsuperscript{13}This suggestion was made by David Freedman.
females operates much as they did in the LAPD patrolled census tracts, and sex misdemeanors operates like it did in the 46 high-risk LAPD patrolled census tracts. In short, sexual orientation, drug use and prostitution are again useful risk variables. The key substantive findings are essentially replicated in a new dataset. At the same time, however, passing this easier test is no substitute for passing the harder test.

V. Discussion and Conclusions

The results suggest that in Los Angeles County it is entirely feasible to collect census tract data on AIDS, crime and the usual census variables. There also does not seem to be anything in the results to date undercutting the paper's substantive aspirations. Indeed, measures of street crime, especially drug use, are apparently useful risk variables at the census tract level. But clearly, most of the work lies ahead.

Perhaps most important, the data need to be updated as more AIDS and crime statistics become available over the next several years. The major costs will stem from the likely need to enter the crime data by hand, coupled with the ongoing problem of collapsing data from reporting districts into census tracts. But once those data are available, it will not only be possible to understand far better how street crime may affect the spread of AIDS, but also how AIDS may affect street crime. It is perhaps in the feedback effects that the most interesting story will be found.
References


