USING POLICE ARREST DATA TO ESTIMATE THE NUMBER OF BURGLARS OPERATING IN A SUBURBAN COUNTY*

Lucius J. Riccio
Police Foundation
Washington, D.C.

Robert Finkelstein
Mitre Corporation
McLean, Virginia

June 1977

* The authors would like to thank the Montgomery County Police Department for their generous cooperation and assistance in this effort, and Larry Siegel for his helpful comments on this paper.

"Using Police Data to Estimate the Number of Burglars Operating in a Suburban County"

Errata

page	line	Comment
7	6	Delete (to be discussed below)
	10	"offenses" should be changed to "detected lapses." Following that sentence, this sentence should be added: A detected lapse is one for which an arrest was made or an exceptional clearance was declared.
9		X^2 should be χ^2 J = .436 should be λ_J = .436 A = .510 should be λ_A = .510 J+A = .440 should be λ_{J+A} = .440 Juvenile + Adult: Accept at $X^2_{.02,1}$
11		Delete (J+A = 100) from last column heading. In last row 46.6 should be 43.6
13		2179 should be 2169 1037 should be 1027

INTRODUCTION

Beginning in the late sixties and intensifying in the early seventies, a strong scholarly interest in examining crime control strategies emerged within the research community. One approach taken to analyze different strategies has been to use systems simulation or stochastic processes modeling to estimate changes in the amount of crime that would result from changes in certain system parameters, such as the probability of arrest or conviction, the activity rate of criminals, or drug program effectiveness. A number of such models have been built, and one in particular has generated considerable interest (Shinnar and Shinnar)[1]. These models have proved helpful in organizing our knowledge about crime control, in testing a variety of strategies, and in providing some guidance about what future research should be undertaken.

Unfortunately, a number of the key variables used in these models cannot be measured directly. For example, we have no way of directly determining the number of active criminals, the percentage of the criminal population that is apprehended by the police, or the activity rate (crimes per year) of those people participating in criminal activity. As such, estimation techniques often must be employed to arrive at reasonable figures for testing crime control strategies.

In addition, accurate estimates for these characteristics of the criminal population would serve as a means of understanding and monitoring the specific crime situation a locale, or the nation as a whole, faces at a particular time. For example, was the rise in crime in the country during the late sixties and early seventies a result of a greater number of criminals participating in illegal behavior or a result of a constant criminal population becoming more active? What was happening to the ability of the police to capture criminals? Was a greater proportion of the criminal population going uncaught? Such information could serve as a real basis for the development of sound crime control strategies as well as vastly improve the quality of research that could be performed on understanding the nature of crime in our society.

A number of analysts have turned their attention to the statistical and mathematical problems associated with estimating some of these key parameters.[2] Willmer (1970) has developed a means of estimating various aspects of the criminal population based on a stochastic model of criminal behavior which uses police arrest data as input. His model has the capability of estimating the size of the criminal population, the proportion of that population apprehended by the police, and the activity rate of both those criminals apprehended by the police and those that are not.

The purpose of this study was to use Willmer's model to estimate the number of burglars in Montgomery County, Maryland, in 1976. This paper discusses the basic form of the model, the data used to calculate the parameters of interest, the implications of the results, and the shortcomings of this particular approach to the problem.

WILLMER'S MODEL

Willmer's model is based on what he calls an "opportunist" or "many temptation" theory of criminal behavior. That is, it is assumed that "the criminal in any given time interval, is subject to a large number of independent temptations or opportunities and that the probability of a lapse* on any particular occasion is small, such that the expected number of lapses has a finite limit in a given period of time."[3] This assumption leads directly to the analytical result that the probability of a criminal lapsing any given number of times in a time interval follows the mathematically well-known Poisson distribution. The mathematics of the model is based on the statistical properties of that distribution.

Intuitively, the "many temptation" theory would seem quite reasonable for certain crimes such as larceny, robbery, auto theft, and burglary, but might not be reasonable for other types such as gambling, prostitution, and narcotics use. Although it does seem reasonable for some types of crimes, there are still a number of problems with accepting such a theory. First, there are some definitional difficulties associated with classifying someone as a criminal as opposed to a noncriminal, especially when one is trying to estimate the number of criminals that participate in illegal activity but go uncaught. The Poisson distribution will yield a data point for the number of "criminals" that commit no detected crimes during a specified time interval. This could occur when these criminals commit crimes that are not reported or not perceived, or if during the time interval of analysis, these criminals do not succumb to temptation. The latter is more fitting of the Poisson assumption. For the

---

* A lapse is a person-crime. Three lapses, for example, may be generated by one person committing three crimes, or three people committing one crime.

sake of theory, it has to be assumed that there is a difference between these non-active "criminals" and law-abiding citizens who commit no crimes and have no criminal tendencies. Definitionally, this can be "resolved" by assuming that the non-active criminals either have, in the past, or will at some time in the future, commit a crime of the type studied.

Second, there are measurement problems since not all crimes are reported to, and therefore, not recorded by the police. The extent to which this disturbs the analysis is not fully known.

Third, using arrest data as a "window" for viewing (in this case, measuring) the criminal population is still a questionable operation. It is not known to what extent the arrest process is truly reactive to criminal behavior. Some recent data indicates that the arrested population of armed robbers has virtually the same demographic characteristics as that of the population of armed robbers as reported by victims in victimization surveys (Hindelang, 1976, pg. 195). That would indicate that arrest is a random sampling process and, therefore, is (from a mathematical point of view) a useful measure for the type of analysis Willmer suggests. However, other data indicates that there are at least two different groups of criminals having widely different personal crime rates and probabilities of arrest (Greenwood, 1977). In addition, the advent and wide-spread use of decoy patrol units might change arrest probabilities and, as such, might alter the utility of arrest data for Willmer's model.

Fourth, it is also assumed that an arrested person is, indeed, the offender sought.

For the purpose of exploring the use of this type of model, it is assumed here that arrest data is a valid means of analyzing the criminal population and that Willmer's assumptions are fundamentally correct. The approach taken is to assume that the model is a good first step and that extensions and additions to the model will improve its validity and utility.

The following equations are pertinent to the model:

(1) $P(x) = \frac{m^x e^{-m}}{x!}$

(2) $N(x) = (n) \frac{m^x e^{-m}}{x!}$

(3) $N(y) = (n) \frac{\lambda^y e^{-\lambda}}{y!}$

(4) $\lambda = pm$

(5) $p = \frac{A}{B}$

(6) $ND = (n)(1 - e^{-\lambda})$

(7) $NT = (n)(1 - e^{-m})$

(8) $NU = n(e^{-\lambda} - e^{-m})$

Where:

P(x) = Probability of an offender committing x (x = 1, 2, 3...) lapses in a time interval.

M = Mean number of lapses in a time interval.

N(x) = Number of offenders who commit x lapses in a time interval.

n = Number of active and "potential offenders.

N(y) = Number of offenders who will be detected for y (y = 1, 2, 3...) lapses in a time interval.

λ = Mean number of detected lapses in a time interval.

p = Probability of being detected per lapse.

A = Number of detected lapses.

B = Total number of lapses

ND = Number of detected offenders.

NT = Total number of offenders.

NU = Number of undetected offenders.

The model can be used to estimate the variables of interest (i.e., the number of offenders, etc.) by taking field data and fitting it to the assumed Poisson distribution. The data can provide an estimate of λ and n (n includes the fitted value for the number of "offenders" having no undetected lapses). With those values plus the number of crimes reported to the police, the analysis can be performed.

THE DATA

Montgomery County, Maryland, is a suburb of Washington, D.C. covering 526 square miles with a population of 585,000. In recent years, it has been listed among the top three wealthiest counties in the United States. The jurisdiction of the county police is the entire county. More than half of all arrests for Part I offenses were juveniles (under 18 years of age).[4]

The county does not tabulate police data in a form suitable for this study: it was necessary to access the original data. Data was collected for 1976. The incident reports consisted of one or more sheets of 8 1/2 x 11 inch forms stapled together containing such information as the name of the victim, location of the crime, a description of the crime, etc. Each report contained information concerning the status of the case: either it was open, or closed by arrest, exception, or as unfounded. In processing the forms for this study, the open cases were counted, but otherwise ignored. The unfounded cases were also simply counted (they consisted of about one percent of the cases). The information extracted from the cases closed by arrest or exception[5] included the name of the suspect, the serial number of the report, the classification of the offense (to be discussed below), and whether the offender was a juvenile or adult.

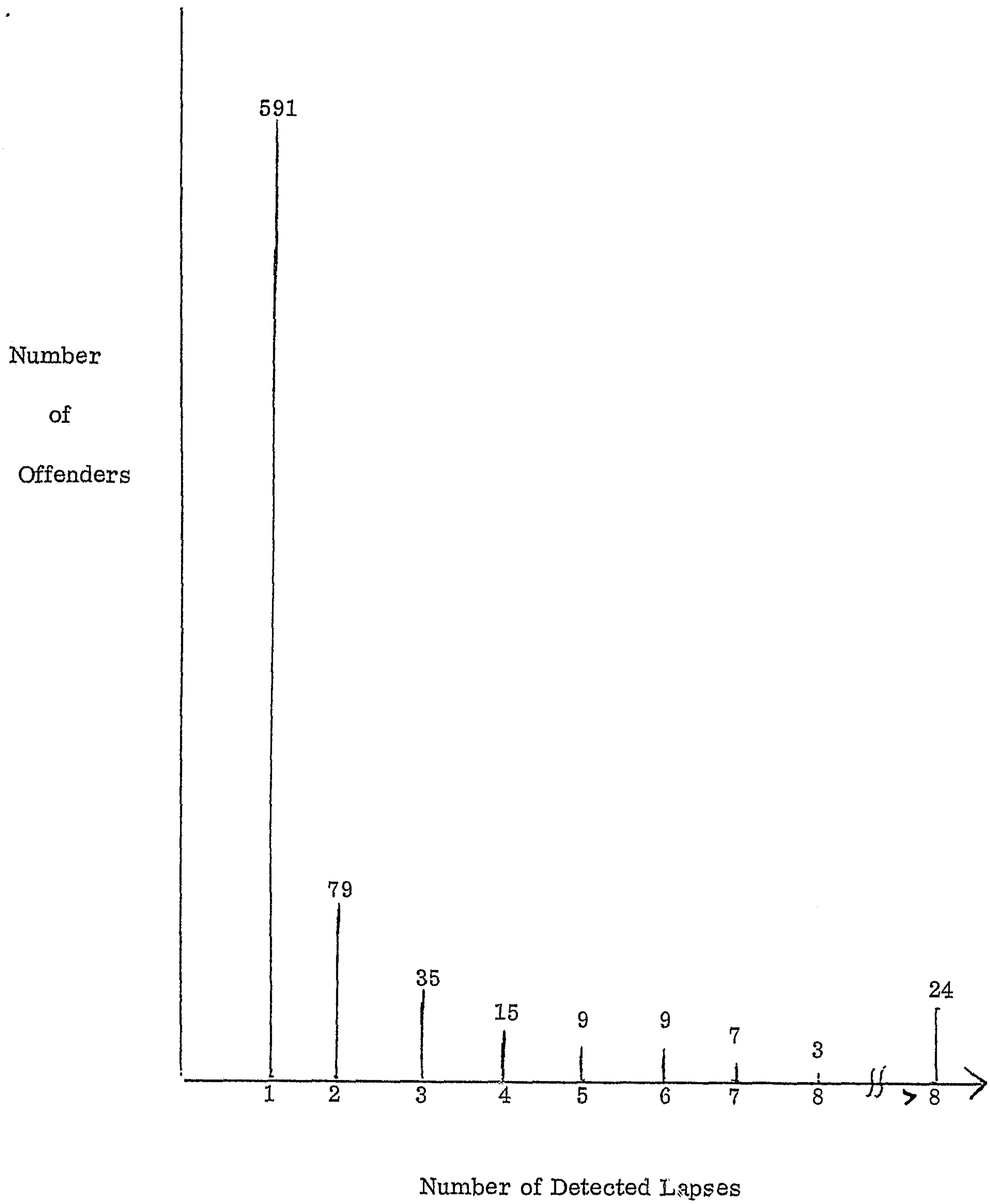
The name of the offender and offense information were placed on 3 x 5 inch cards, one card per offense, for ease in processing. The cards were later alphabetized by name so that the number of offenses for each person could be tabulated.

ANALYSIS

Figure 1 is a frequency diagram of the number of offenders committing x number of detected lapses for the sum of adult and juvenile offenders. Due to the unexpectedly large number of high-activity offenders, fitting the data to a Poisson distribution became a somewhat difficult task. As such, the data was divided into two regimes: the P-regime containing offenders with five or fewer detected lapses and the D-regime containing offenders with more than five detected lapses.

Table 1 summarizes the number of detected offenders and lapses for each category. Table 2 provides the number of lapses and offenders for each category and regime combination and the ratio of lapses to offenders. Table 3 provides the percentage of offenders in each category; the percentage of offenders in each category relative to the total offenders in each regime; the percentage of lapses relative to each

Figure 1. -- Distribution of Offenders Having a Given Number of Detected Lapses for the Sum of Adult and Juvenile Offenders



category and each regime. It can be seen in Table 3 that well over 90% of all categories of detected offenders fall in the P-regime, but that these offenders account for a bit more than 50% of the detected lapses.

Before fitting Poisson distributions to the offenders versus offense frequency tabulations, all offenders who committed more than five lapses during 1976 were placed in the D-regime. The Poisson parameters obtained for the juvenile (J), adult (A), and juvenile plus adult (J+A) populations were:

J = .436

A = .510

J+A = .440

A χ^2 test was performed for each fitted Poisson with the following results:

Juvenile	: Accept at $\chi^2_{.05,1}$
Adult	: Accept at $\chi^2_{.05,1}$
Juvenile + Adult	: Accept at $\chi^2_{.05,1}$

To relate lapses to crimes, the total number of lapses was divided by the total number of crimes (burglaries)[6] as follows:

$$\frac{\text{number of lapses cleared}}{\text{number of crimes cleared}} = \frac{1699}{1557} = 1.09 \frac{\text{lapses}}{\text{crime}}$$

Because there was no distinction between adult crimes and juvenile crimes, the ratio is valid only for the sum of the juvenile and adult populations. The remaining analysis will be concerned only with the sum of the juvenile and adult populations.

TABLE 1

SUMMARY OF DETECTED OFFENDERS AND LAPSES

CATEGORY	TOTAL OFFENDERS	TOTAL LAPSES
Juvenile	486	895
Adult	286	804
Adult + Juvenile	772	1699

TABLE 2

CATEGORY AND REGIME DETECTED OFFENDERS AND LAPSES

CATEGORY	REGIME	NUMBER OF OFFENDERS	# LAPSES	RATIO
Juvenile	P	467	596	1.3
Adult	P	262	363	1.4
Adult + Juvenile	P	729	959	1.3
Juvenile	D	19	299	15.7
Adult	D	24	441	18.4
Adult + Juvenile	D	43	740	17.2

TABLE 3

CATEGORY AND REGIME PERCENTAGES

CATEGORY	REGIME	PERCENTAGE OFFENDERS WITHIN CATEGORY	PERCENTAGE OFFENDERS WITHIN REGIME	PERCENTAGE LAPSES WITHIN CATEGORY	PERCENTAGE LAPSES (J+A=100) WITHIN REGIME
Juvenile	P	96.1	64.1	66.6	62.1
Adult	P	91.6	35.9	45.2	37.8
Juvenile + Adult	P	94.4	100.0	56.4	100.0
Juvenile	D	3.91	44.2	33.4	40.4
Adult	D	8.39	55.8	54.8	59.6
Juvenile + Adult	D	5.57	100.0	46.6	100.0

Using equation 5[7],

$$P = \frac{(959 \text{ lapses in P-Regime for J+A population})}{(5349 \text{ crimes})(1.09 \text{ lapses/crime})(.564 \text{ fraction of lapses in P-Regime})}$$

$P = .292$

From the fitted Poisson distribution,

$\lambda = .440$

$n = 2179$[8]

From equation 4,

$$m = \frac{\lambda}{p} = \frac{.440}{.292} = 1.51$$

Using equation 8, the number of uncaught P-Regime burglars is given by

$$NU = 2179\ (e^{-.44} - e^{-1.51}) = 922$$

An estimate for the D-Regime uncaught burglars may be obtained as follows:

$$NU' = \Big[(5349 \text{ crimes})(1.09 \text{ lapses/crime})\left(\frac{1}{17.2} \text{ offenders/lapse}\right)(.436 \text{ fraction of lapses in D-Regime})\Big] - 43 \text{ detected offenders}$$

$NU' = 105$

Total uncaught burglars = 922 + 105 = 1027

The uncaught burglars represent

$$\frac{1027}{1027 + 729 + 43} = 0.57$$

or 57% of the total population of burglars.[9] They committed 71% of the crimes:

$$\frac{3792}{5349} \times 100 = 70.9 \text{ percent}$$

In other words, the Montgomery County Police captured 43% of the burglar population at one time or another over the course of 1976. The captured burglars accounted for 29% of the lapses. Of the 2179 active and "potential" burglars, 772 were captured by the police, 1037 committed burglaries but were not captured, and 370 people did not commit any burglaries during 1976 but could be viewed as "potential" burglars.

CONCLUSIONS

This paper has been an attempt to use Willmer's model for estimating certain characteristics of the criminal population. It was not an attempt to validate his model. It cannot be said with confidence at this point, that the figures calculated here are accurate. The purpose was to apply Willmer's model to demonstrate its application and to learn more about the utility of the model.

Althought it is virtually impossible to validate such a model, little was learned in our experience to indicate that the mathematics are not a useful first step in estimating important nondirectly-measureable aspects of the criminal population. Improvements certainly could be made by performing victimization studies and by delineating different segments of the criminal population as well as by detailing and strengthening the assumption base of the model.

The figures calculated seem to be very sensitive to small changes in the calculation of certain parameters. It has been pointed out that if the ratio of lapses to crimes differs somewhat for the burglars in the two statistical regimes (it was assumed they were the same), the ratio of caught to uncaught burglars changes substantially.

There is one aspect of the mathematics that needs more analytical attention. The calculations presented estimate that although the police apprehend a large share of the burglar population, the group they arrest accounts for a less than proportional share of the burglaries. For this relationship to change (i.e., for the data to indicate that the police capture those criminals who account for more than a proportional share of the burglaries), the Poisson distribution for detected lapses would have to have a higher mean, a higher value for λ. That is, graphically, the plot of the distribution would have to shift to the right. What that means is that the police would have to capture these high activity people more often. However, the arrest of a burglar might significantly change that individual's personal crime rate. That is, if arrested and incarcerated, that person cannot commit any additional crimes until he/she is released. As such, it is possible that the estimate is biased toward the low side. Of course, the bias might not be significant because as recent data indicates, very few arrests lead to a significant amount of incapacitation (PROMIS, 1977). On the other hand, the data might always show this relationship since, by definition, the criminals who are committing the burglaries are those whom you do not capture. That is, you can't capture this year's high activity criminals because when you do capture one, he/she is no longer a high activity criminal.

As a final note, the authors feel there is great utility in developing estimation models of this type. Willmer's approach appears to have an appropriate mathematical foundation. Extensions of his work should be attempted.

NOTES

1. For example, see Abraham (1972), Aerospace Corporation (1974), Ari-Itzhak and Shinnar (1973), Fey and Wadsworth (1973), Grieco (1974), Hirsch et al, (1971), Rardin and Gray (1973), and Riccio (1971).

2. See Nagin and Blumstein (1975) and Blumstein and Greene (1976), as well as Willmer (1970) and (1972).

3. Willmer (1970) pg. 101

4. "Annual Report 1975" (Montgomery County, Maryland: Department of Police 1975), p.55.

5. Closure by exception includes those cases where the police make no arrests even though they believe they know the identity of the offenders. Reasons for not arresting the suspect(s) include: insufficient evidence for an indictment; lack of necessity because the suspects have been arrested or are in prison on a related charge; lack of desire to press charges by the victim or police.

6. "Department of Police Monthly Report" (Montgomery County, Maryland: Department of Police, December 1976), provides the following: Number of burglaries = 5349; Number of burglaries cleared = 1557.

7. In estimating p, it is assumed that the ratio of lapses to crimes is independent of whether or not the crimes are cleared.

8. This includes an estimated 1407 "potential" offenders that committed no crimes in 1976 or offenders who went undetected.

9. This compares with the 56% obtained by Willmer for 1964 in the English Bedford Division. Other years ranged from 41 to 47 percent (Willmer, p. 11).

REFERENCES

1. Abraham, S.C. "Simulation Modelling in Criminal Justice: An Approach and Demonstration Using System Dynamics." Graduate School of Management, University of California at Los Angeles, July, 1972.

2. Aerospace Corporation. Strategic Planning. Washington, D.C., July 1, 1974.

3. "Annual Report-1975" Department of Police, Montgomery County, Maryland.

4. Ari-Itzhak, Benjamin and Shinnar, Reuel. "Quantitative Models in Crime Control." Journal of Criminal Justice, 1973, 1 : 185-217.

5. Blumstein, A. and Greene, M. "Estimation of the Offender Crime Rate." ORSA/TIMS Joint National Meeting, Miami, Florida, 1976.

6. Fey, W.R. and Wadsworth, H.M. "Criminal Justice Model" Georgia Institute of Technology, February 17, 1973.

7. Greenwood, Peter and Joan Petersilia, "Alternative Policy Options for Dealing With the Serious Habitual Offender," Rand Corporation, Santa Monica, California 90406.

8. Grieco, Donald M. "Minimizing the Crime Rate Through the Allocation of Uniform and Plainclothes Patrol Officers." ORSA/TIMS National Meeting, Fall 1974.

9. Hindelang, Michael J. Criminal Victimization in Eight American Cities. Ballinger Publishing Company. Cambridge, Massachusetts. 1976.

10. Hirsch, Gary, Roberts, Edward, and Gilbert Levin, "Narcotics and the Community: A System Simulation," Pugh-Roberts Associates, Cambridge, Mass., 1971.

11. "Monthly Report - December, 1976" Department of Police, Montgomery County, Maryland.

12. Nagin, D. and Blumstein, A. "Analysis of Arrest Rates for Trends in Criminality." Journal of Socio-Economic Planning Sciences, September, 1975.

13. PROMIS, "Highlights of Interim Findings and Implications," Institute for Law and Social Research, 1125 Fifteenth Street, N.W., Washington, D.C. 20005

14. Rardin, Ronald and Paul Gray. "Analysis of Crime Control Strategies," Journal of Criminal Justice, Vol. 1, (1973), p.339.

15. Riccio, Lucius J. "A Model for the Study of Crime Control Administration," May, 1971.

16. Shinnar, S. and Shinnar, R. "The Effects of the Criminal Justice System on the Control of Crime," City College of New York, 1975.

17. Willmer, M.A.P. Crime and Information Theory. Edinburgh, U.K.: Edinburgh University Press, 1970.

18. Willmer, M.A.P. "On the Estimation of the Number of Persons Entering and Leaving Crime Annually" Journal of Crime and Delinquency, 1972, p. 175.

END