PROJECTED PERCENTAGE OF U.S. POPULATION WITH CRIMINAL ARREST AND CONVICTION RECORDS

by Ronald Christensen

Contents

The object of this note is to provide an estimate of the percentage of the future U.S. population which will have a criminal arrest record resulting from at least one non-traffic arrest, and of the percentage which will have a criminal conviction record resulting from such an arrest. Making such estimates requires assumptions about how trends will develop in the future. In general, we make the assumption of "steady state," or that the current situation will continue into the future. This is not intended to suggest that the current situation will necessarily continue, but that the projections will be as indicated if it does continue. If, for instance, the projection of arrest probability is viewed as undesirably high, then that would suggest a reconsideration of the factors making it so high. The value of such projections lies in stimulating such considerations much more than in a literal prediction of the future.

Making these projections also requires various data on such parameters as number of arrests, population distribution, and the virgin arrest ratio (i.e., the portion of arrestees who have never before been arrested). Some of these parameters, especially the virgin arrest ratio, are difficult to estimate accurately from available data. Wherever possible, conservative estimates have been used. Since the validity of any analyses such as those reported here are inherently limited by the available data, we hope that this paper will stimulate the collection of more accurate and complete data.

ARRESTS

The arrest data [1] used are based on crimes reported in the FBI's Uniform Crime Reports (UCR) for 1965. These crimes are listed together with arrest totals for the reporting agencies:

Offense charged	All ages	Ages under 18	Ages over 18
Total	5,031,393	1,074,485	3,956,908
Criminal homicide:			
Murder and nonnegligent manslaughter	7,348	635	6,713
Manslaughter by negligence	2,815	196	2,619
Forcible rape	10,734	2,245	8,489
Robbery	45,872	13,813	32,059
Aggravated assault	84,411	12,950	71,461
Burglary, breaking or entering	197,627	102,472	95,155
Larceny, theft	383,726	210,469	173,257
Auto theft	101,763	63,596	38,167
Total of above offenses	834,296	406,376	427,920
Other assaults	207,615	31,948	175,667
Arson	6,187	4,031	2,156
Forgery and counterfeiting	30,617	2,962	27,655
Fraud	52,007	1,796	50,211
Embezzlement	7,674	275	7,399
Stolen property: buying, receiving, possessing	19,060	6,720	12,340
Vandalism	89,668	68,785	20,883
Weapons: carrying, possessing, etc	53,585	10,985	42,600
Prostitution and commercialized vice	33,987	839	33,148
Sex offenses (except forcible rape and prostitution)	58,205	14,097	44,108
Narcotic drug laws	46,069	5,345	40,724
Gambling	114,294	2,561	111,733
Offenses against family and children	60,981	648	60,333
Driving under the influence	241,511	1,937	239,574
Liquor laws	179,219	48,456	130,763
Drunkenness	1,535,040	25,912	1,509,128
Disorderly conduct	570,122	93,472	476,650
Vagrancy	120,416	7,894	112,522
All other offenses (except traffic)	531,970	156,310	375,660
Suspicion	76,346	20,612	55,734
Curfew and loitering law violations	72,243	72,243	----------
Runaways	90,281	90,281	----------

During the last 5 years, the probability of being arrested as a function of age has been increasing for each age group of the population, the increase being most pronounced for the younger age groups. However, for purposes of this calculation, we will make the conservative assumption that the age-specific arrest probabilities for future years will remain the same as they were in 1965. This tends to underestimate the probability of eventual arrest. Yet, this probability will be found to be strikingly high even with this and other conservative assumptions.

[1] "Crime in the United States, Uniform Crime Reports—1965"; July 28, 1966, p. 112. Covers 4,062 reporting agencies which represented an estimated 1965 population of 134,095,000.

METHOD USED

Suppose P_t is the probability of a person in the class of interest being arrested as a t-year-old. Then the probability of an individual in this class being arrested during his life is

$$P = P_0 + (1-P_0)P_1 + (1-P_0)(1-P_1)P_2 + \ldots$$

$$= \sum_{t=0}^{T} (1-P_0) \ldots (1-P_{t-1})P_t$$

$$= 1 - \prod_{t=0}^{T} (1-P_t)$$

where T is a large number, say 90 years, after which age few people are arrested for the first time.

However, suppose p_t is the probability of an individual in the class of interest being arrested *for the first time* as a t-year-old. Then

$$P = \sum_{t=0}^{T} p_t$$

The two equations are, of course, equivalent since

$$p_t = (1-P_0) \ldots (1-P_{t-1})P_t$$

Our model is steady state in the sense that we assume that the age-dependent arrest probabilities for 1965 carry forward unchanging into the future.

In figure J–1, those persons 5 years old in 1965 fall in the lower left-hand shaded box in the figure. (We use 5-year-olds since this is approximately the age at which some arrests begin to occur.) As time progresses, this group advances up the series of diagonal boxes. Because of the steady state assumption, all boxes along a horizontal line represent identical arrest probabilities. Thus, we refer back across to the associated box in the 1965 data base to establish the appropriate p_t.

Let V_t be the number of first arrests in the t^{th} box and M_t be the total number of people in the t^{th} box. Then

$$p(t) = \frac{V_t}{M_t}$$

is the probability of first arrest for someone in the t^{th} box.[2]

The proportion of the 5-year-olds who, from survival statistics, will live to be t years old is L_t (see table J–1). Thus, the probability of a person who is 5 years old today being first arrested as a t-year-old is $p_t = L_t p(t)$.

Summing, we get:

$$\sum_{\tau=5}^{t} p_\tau$$

as the probability of someone 5 years old today being arrested by the time he is a t-year-old. The summation is proper since being arrested for the first time at one age and being arrested for the first time at a different age are disjoint events.

FIGURE J-1. THE POPULATION MODEL

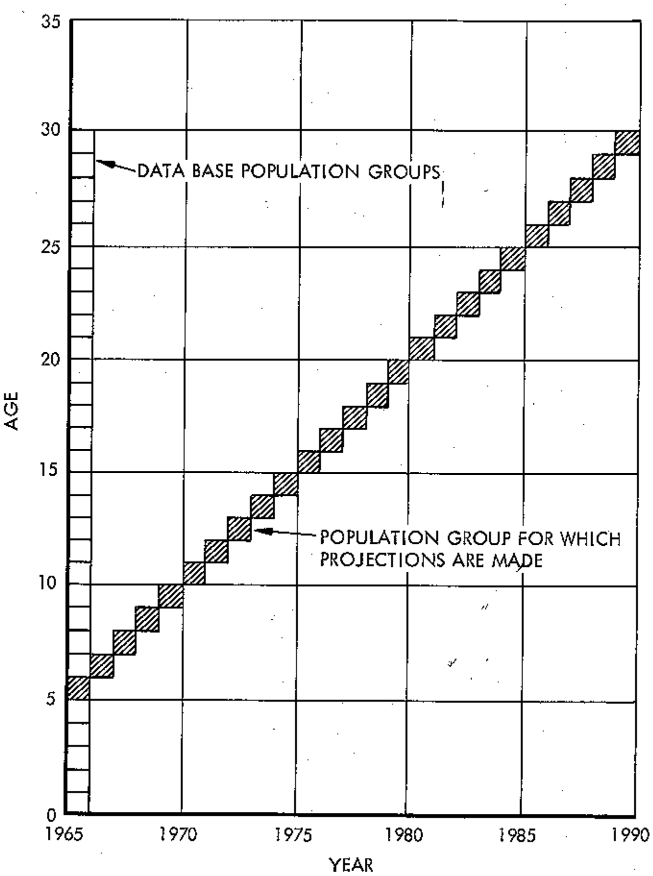


Table J–1.—Probability of a 5-Year-Old Surviving to Age t

Age (t)	Probability (L_t) of survival	Age (t)	Probability (L_t) of survival
5	1.0000	40	0.9578
10	.9978	45	.9403
15	.9957	50	.9135
20	.9905	55	.8727
25	.9843	60	.8143
30	.9777	65	.7341
35	.9694	70	.6266

Source: 1966 Statistical Abstract.

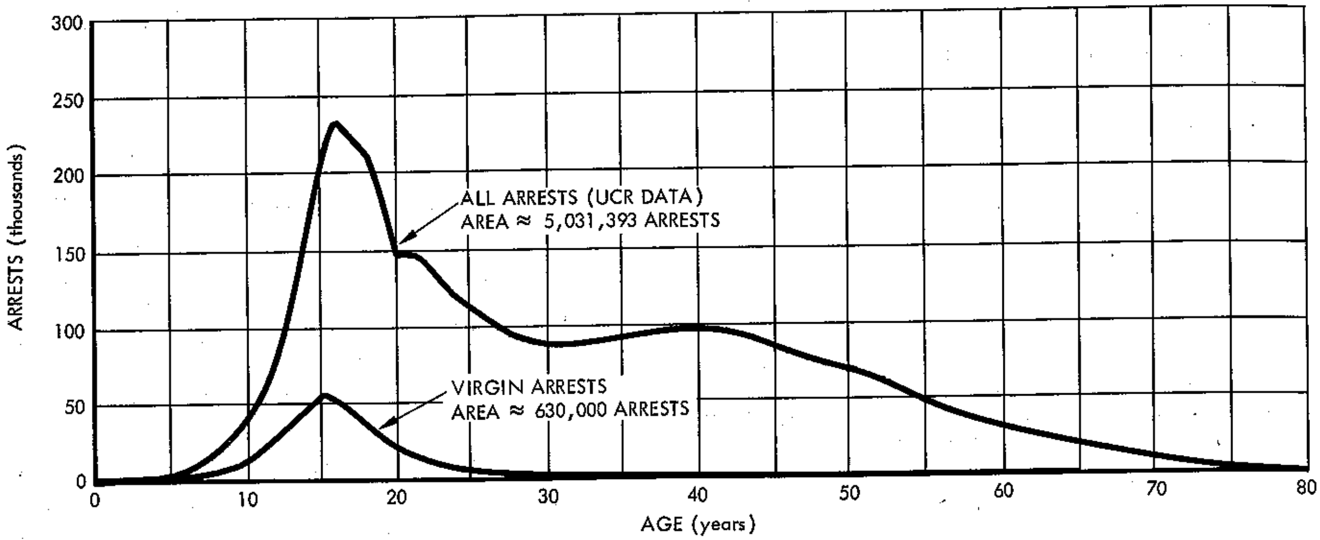
DATA ON ARRESTS AND VIRGIN ARREST RATIO

Figure J–2 shows the distribution of the total number of arrests in 1965 by age,[3] as recorded in the 1965 UCR. The total area below the top curve integrates to a total of 5,031,393 arrests by reporting agencies servicing a population of 134,095,000 (about 69 percent of the total census estimate of 194,400,000 for 1965). We will assume that the crime characteristics of this population

[2] The denominator must be the total number of people in the t^{th} box, and not just those never arrested at a younger age, in order to obtain the cumulative probability by simple summation as we do below. If we reduce the denominator as stated, we get the conditional probability of a t-year-old being arrested given that he has never been arrested. The ratio V_t/M_t, on the other hand, is the joint probability of a t-year-old being arrested and never having been arrested before. Not only would using the conditional probability rather than the joint probability lead to a more cumbersome formulation, it would require data presently unavailable; namely, data on the arrest history of the population as a whole as well as on those arrested in 1965.

[3] An interesting anomaly is the unusual secondary peak in the rates of arrest for various crimes of reported 18-year-olds and the corresponding depressions for reported 17-, 19-, and 20-year-olds in fig. J–2. (The peak is even more pronounced for graphs drawn by crime type.) It remains to be seen whether this is a real phenomenon or a peculiarity in reporting the age of arrestees.

FIGURE J-2. 1965 ARRESTS BY AGE FOR ALL NONTRAFFIC OFFENSES



are representative of the United States in general, although we will later correct for the relative disproportion of urban coverage by the reporting agencies.

The next problem is to find the fraction of all arrests which are virgin arrests; i.e., arrests of individuals who have never been arrested before.[4] This datum is crucial to the calculation, and has proved most difficult to come by. Almost all potential sources for this datum proved to be insufficiently complete, resulting in an unrealistically high figure. For example, according to the Bureau of Criminal Statistics for California, about 32 percent of the arrests in San Mateo County between July 1, 1961, and June 30, 1962, were of individuals with no prior record.[5] A breakout of this percentage by crime type is given as follows:

Felonies	Virgin Percent	Misdemeanors	Virgin Percent
Homicide	25.0	Assault	38.9
Robbery	14.8	Petty theft	50.0
Aggravated assault	49.1	Drunk driving	36.5
Burglary	25.8	Checks	38.9
Grand theft, except auto	41.7	Drunk	22.5
Auto theft	23.7	Disturbing peace	35.4
Forgery of checks	35.3	Malicious mischief	50.0
Rape	40.0	Sex offenses	49.0
Other sex	66.7	Other misdemeanor	36.0
Narcotics	10.0		
Other felony	24.3		

These included both felony and misdemeanor arrests, and the misdemeanor arrests included roughly the usual proportion of such things as arrests for drunkenness. Aside from the fact that San Mateo is not representative of the nation, the principal reason for the extremely high figure is that juvenile arrest was not included either as a criterion for inclusion in the sample or as a part of "prior record" in the data. It covered only adult records of adults arrested during the period. Almost one-quarter of all arrests are of persons under 18 years of age.

The FBI criminal history file,[6] which yields a virgin arrest ratio of about one-quarter, is also not applicable to this problem. First, it includes only fingerprintable offenses (even not all of these), and therefore excludes drunkenness, etc., from almost all jurisdictions. Second, it is a file of certain arrestees rather than a representative sample of all arrests during some given period. Similar remarks apply to State criminal history files, such as the New York State file which, by counting virgin arrests in arrest reports received, was found[7] to yield a ratio of about one-third for the year 1965.

On September 22, 1966, a random sample of arrest records was taken from the District of Columbia Metropolitan Police Communications and Records Bureau criminal history file. The fraction of these arrests which were of individuals with no prior record in the file is shown in figure J-3 as a function of the year of the arrest. The downward trend seems to suggest that the file has been becoming more complete over the last 5 years (although it is probably also due in part to the fact that the percentage of the population with no record has been decreasing as the arrest rate has been increasing). Here again we get a ratio somewhere between one-quarter and one-third due to such factors as the absence of juvenile records in the file and the absence of records from other jurisdictions.

The only source of data which proved sufficiently complete was a study of 1965 juvenile referrals and adult arrests in the District of Columbia based on probation de-

[4] In the steady state model (but allowing arbitrary population-age distribution), the virgin arrest ratio r is exactly the inverse of the average number of arrests a during the lives of those arrested at least once:

$$r = \frac{\text{number of virgin arrests in any year}}{\text{total number of arrests in the year}} = \frac{\text{number of people arrested in a long-time period}}{\text{total number of arrests in the long-time period}} = \frac{1}{a}$$

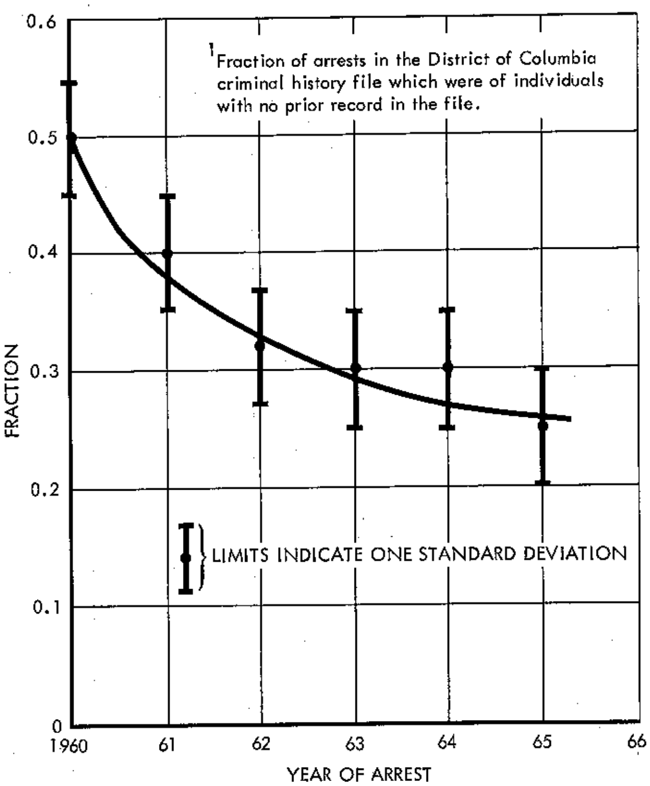
For example, if the average arrest career is 8 arrests, then on the average ⅛ of all arrests in any given time period are of persons never arrested before.

[5] Bureau of Criminal Statistics, California Department of Justice, "San Mateo County Statistical Reporting Project," 2700 Meadowview Rd., Sacramento, Calif. For the period 1962–63, the number was about 29 percent.

[6] UCR, 1965, p. 28.

[7] Private communication, September 1966.

FIGURE J-3. FRACTION OF VIRGIN ARRESTS IN THE DISTRICT OF COLUMBIA CRIMINAL HISTORY FILE [1]



partment investigations.[8] For juveniles, it was found that about 39 percent of those referred to juvenile court had no previous referrals.[9] For adults, it was found that about 15 percent of the white arrestees and 8 percent of the Negro arrestees who were convicted of a felony in 1965 had no prior arrest record for either a misdemeanor or a felony.[10] These last figures are probably lower than the national averages both because District of Columbia, being urban, is not a truly representative area, and because this sample is probably biased toward the repeater, being of convicted félons only rather than all arrestees.

According to the UCR, 29 percent of the U.S. arrestees in 1965 were Negroes, while 71 percent were whites and others. Thus, projecting the above fractions onto the nation as a whole, we obtain:

$$(0.71)(0.15)+(0.29)(0.08)=0.13$$

as an estimate of the fraction of all adults arrested in 1965 who never had a prior arrest.

Now we must deal with the fact that some persons were arrested more than once in 1965. Let us define:

N_A=Number of arrests in 1965
N_I=Number of individuals arrested in 1965
N_V=Number of virgin arrests in 1965

Then, for juveniles:

$$\left(\frac{N_V}{N_A}\right)_{\text{juveniles}}=\frac{\left(\frac{N_V}{N_I}\right)_{\text{juveniles}}}{\left(\frac{N_A}{N_I}\right)_{\text{juveniles}}}\approx 31 \text{ percent}$$

using the figures:

$$\left(\frac{N_V}{N_I}\right)_{\text{juveniles}}\approx 39 \text{ percent}$$

and:

$$\left(\frac{N_A}{N_I}\right)_{\text{juveniles}}\approx {}^{11}1.25$$

For adults:

$$\left(\frac{N_V}{N_A}\right)_{\text{adults}}=\frac{\left(\frac{N_V}{N_I}\right)_{\text{adults}}}{\left(\frac{N_A}{N_I}\right)_{\text{adults}}}\approx 7.6 \text{ percent}$$

using the figures:

$$\left(\frac{N_V}{N_I}\right)_{\text{adults}}\approx 13 \text{ percent}$$

and:

$$\left(\frac{N_A}{N_I}\right)_{\text{adults}}\approx {}^{12}1.7$$

This gives us estimates of the virgin arrest ratio for two different age groups, those below 18 and those 18 and above. (Although the juvenile-adult dividing line varies from jurisdiction to jurisdiction and can even depend upon the case within a jurisdiction, we will use the average of about 18 years of age for purposes of these computations.) However, if we want to know the probability of an individual acquiring a record by the time he reaches a given age, then we must know explicitly how the ratio

[8] Report of the President's Commission on Crime in the District of Columbia; 1966; appendix.

[9] Ibid, p. 490.

[10] Ibid, p. 488.

[11] Based on data from the 52d Annual Report of the County Court of Philadelphia, 1965, pp. 111 and 131.

[13] A conservative estimate of this number can be obtained as follows: Let a= average lifetime number of arrests of persons arrested at least once in lifetime, and T_a=average arrest career (average number of years between first and last arrests of persons arrested at least once). Then the average number of arrests per year during an arrest career is a/Ta. If all we know is that an individual was arrested at least once in 1965, our estimate of the number of additional times that individual was arrested during that year should be no greater than

$$\frac{a-1}{T_a}.$$

Since N_A-N_I equals the number of persons arrested in 1965 times the average number of "additional" arrests per person in 1965, we get an upper bound of

$$\frac{N_A}{N_I}=1+\frac{N_A-N_I}{N_I}$$

$$=1+\frac{N_I\left(\frac{a-1}{T_a}\right)}{N_I}$$

$$=1+\frac{a-1}{T_a}$$

At this point, we imagine leaving a as a parameter, and solving for it by setting $r=1/a$ when we finally compute the overall virgin arrest ratio. Using $T_a\approx 10$ (FBI criminal career data), this gives the equation:

$$1/a=\frac{0.31\,(N_A)_{\text{Juveniles}}+\frac{0.13(N_A)_{\text{Adults}}}{0.9+0.1a}}{(N_A)_{\text{Juveniles}}+(N_A)_{\text{Adults}}}$$

Inserting the 1965 UCR values:

$$(N_A)_{\text{Juveniles}}=1,074,485$$
$$(N_A)_{\text{Adults}}=3,956,908$$

The solution is $a\approx 8$.

This gives us:

$$\left(\frac{N_A}{N_I}\right)_{\text{Adult}}\approx 1.7.$$

of first arrests to all arrests varies with the age of the individual. We know that it is generally a decreasing function of age. But there appears to be no data from which we can directly determine the functional dependence. However, we do have sufficient information to construct a function which will be sufficiently accurate for our purposes. The data points on figure J–4 represent the fraction of juvenile referrals, by age, in the Philadelphia County Court in 1965 with no prior court record. Although these data refer to court records rather than arrest records, they are still useful in approximating the shape of the curve in the 6- to 17-year-old range, and providing a basis for estimating its shape outside this range.

We now seek a curve, representing the ratio r_t as a function of t, of this general shape which satisfies the constraints:

$$\frac{\sum_{t=1}^{17} r_t N_t}{\sum_{t=1}^{17} N_t}=0.31$$

$$\frac{\sum_{t=18}^{90} r_t N_t}{\sum_{t=18}^{90} N_t}=0.076$$

where:

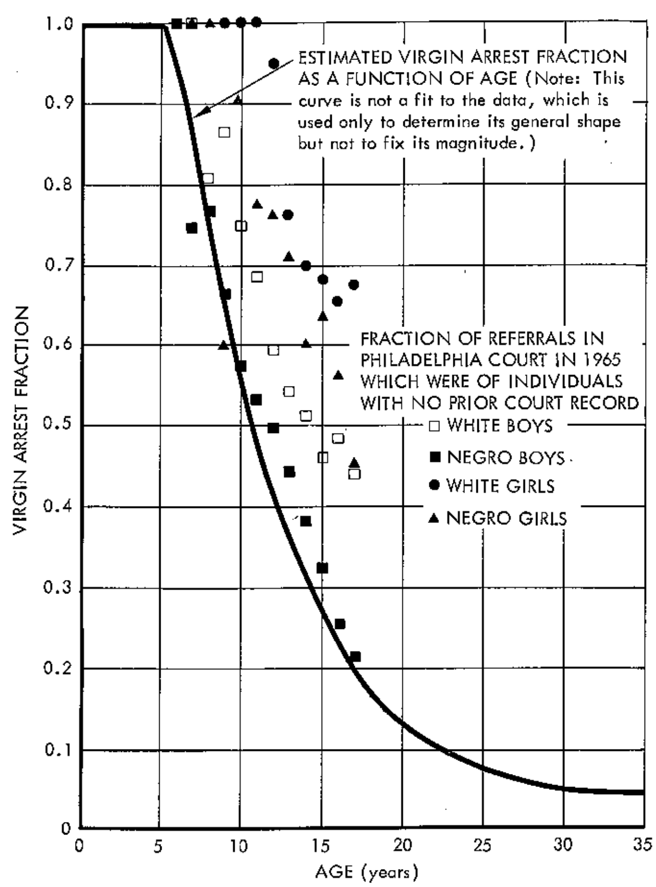
N_t=Number of UCR arrests in 1965 of t-year-olds

A curve which meets these requirements is also shown in figure J–4.

Having found r_t, we can plot $r_t N_t$ as a function of t. This is the lower curve on figure J–2, representing the number of virgin arrests in 1965 as a function of age of individuals arrested. The number of virgin arrests is about 630,000 out of a total of 5,031,393, giving an overall virgin arrest ratio of about one-eighth, or 12.5 percent.

In general the age-specific ratio, r_t, for any given age t, is highest for white females, next highest for Negro females, next for white males, and lowest for Negro males. However the available data is insufficient to enable us to estimate very reliably the difference in the functional form of r_t for each of these four categories. Instead when we come to the end of the calculation we will simply correct the weighted average to 0.15 for whites and 0.08 for Negroes, in accordance with the District of Columbia Crime Commission data. Since males constituted the bulk, 94.1 percent, of the arrestees in that sample, these figures are really estimates for males and not for females. Data for 1965 from the County Court of Philadelphia indicates that the virgin arrest ratio is about two times as great for females as for males. We will also make this correction at the end of the calculation.

FIGURE J-4. VIRGIN ARREST FRACTION AS A FUNCTION OF AGE



ARREST PERCENTAGES

Figure J–5 shows M_t, the number of t-year-olds in the United States in 1965, as a function of age t, the total population being about 194,400,000.

The curve of figure J–6 shows the probability of an individual being arrested during the year for the first time in his life as a function of age. It is calculated by the formula:

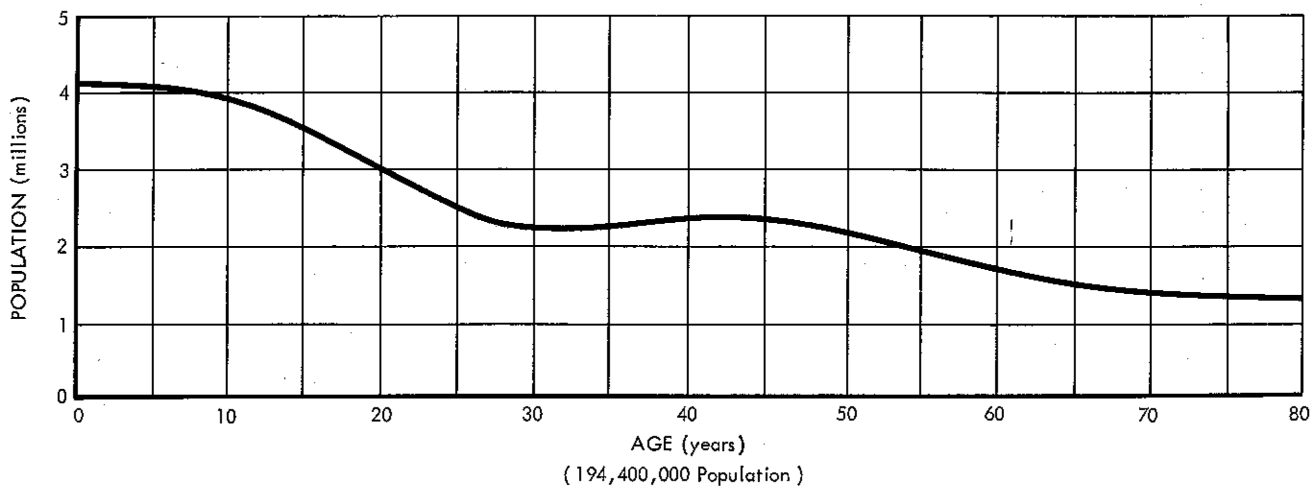
$$p(t)=\left(\frac{\text{UCR first arrests in 1965 of } t\text{-year-olds}}{t\text{-year-olds in United States in 1965}}\right)$$

$$\times\left(\frac{\text{1965 United States population}}{\text{1965 UCR population}}\right)$$

$$=\left(\frac{r_t N_t}{M_t}\right)(1.45)$$

The fraction of total 1965 UCR arrests which were of males was 0.881 and of females was 0.119. Thus, the

FIGURE J-5. 1965 U.S. POPULATION AGE DISTRIBUTION



cumulative probability of the child born in 1965 having an arrest record by the time he is T years old is given by:

Male:

$$P_m(T) \approx \sum_{t=1}^{T} \frac{0.881}{0.492} p_t$$

Female:

$$P_f(T) \approx 2 \sum_{t=1}^{T} \frac{0.119}{0.508} p_t$$

where:

$$p_t = L_t p(t)$$

As mentioned previously the factor of 2, to correct the female virgin arrest ratio, is included in the above expression for females.

These curves are plotted on figure J–7. They show, for example, that the probabilities of eventual arrest by the time of life-expectancy age to be 0.52 for males and 0.13 for females. These figures are not yet corrected for the disproportionate urban coverage of the UCR.

The lifetime arrest figure for males can be checked against intuition very easily. Assume for simplicity that all first arrests occur at age 16. There were about (0.492) (3,480,000) = 1,710,000 16-year-old males in the United States in 1965, and about $\frac{1,710,000}{1.45} = 1,180,000$ 16-year-old males in the population covered by the 1965 UCR arrest data. The 1965 UCR records 4,431,625 male arrests.[13] So with this simple intuitive model, the probability of a male being arrested during his life is about:

$$\frac{\frac{1}{8}(4,431,625)}{1,180,000} \times 100 \text{ percent} \approx 47 \text{ percent}$$

A similar intuitive calculation for females gives 12 percent (using an overall virgin arrest ratio of one-quarter for females).

This approximate calculation also shows that any error in the original data is not compounded into a larger error in the final result. The final result is directly proportional to the virgin arrest ratio, the total number of arrests, and inversely to the population. Thus, a 10-percent error, for example, in any of these data would simply produce a 10-percent error in the final result. These same considerations apply to the more detailed computations for each age group.

The principal reason for the difference in results of the two calculations, 52 and 13 percent versus 47 and 12 percent, is that in 1965 there were fewer persons in each of the age groups beyond age 16, and that the first calculation took into account a significant fraction of the first arrests being distributed over these older age groups. Each of these age groups then produced a contribution with a smaller denominator to the sum. Another reason, somewhat less important numerically, is that the first calculation accounted for the survival probabilities, L_t.

Both of these calculations depend critically upon the somewhat uncertain virgin arrest ratio. The value used here has been based on a rather conservative set of assumptions. Nevertheless, there will always be uncertainty resulting from the problem of incompleteness of arrest records. However, on the basis of the data which was available, it appears safe to conclude that if future arrest rates are as high as those in 1965, then the lifetime arrest probabilities will be at least 40 percent for males and 10 percent for females, and possibly even higher.

Of course, figure J–7 probably overestimates the fraction of the *present* population that has an arrest record (say, by looking at the figures corresponding to the cur-

[13] UCR, p. 115.

FIGURE J-6. 1965 VIRGIN ARREST PROBABILITY BY AGE FOR ALL NONTRAFFIC OFFENSES

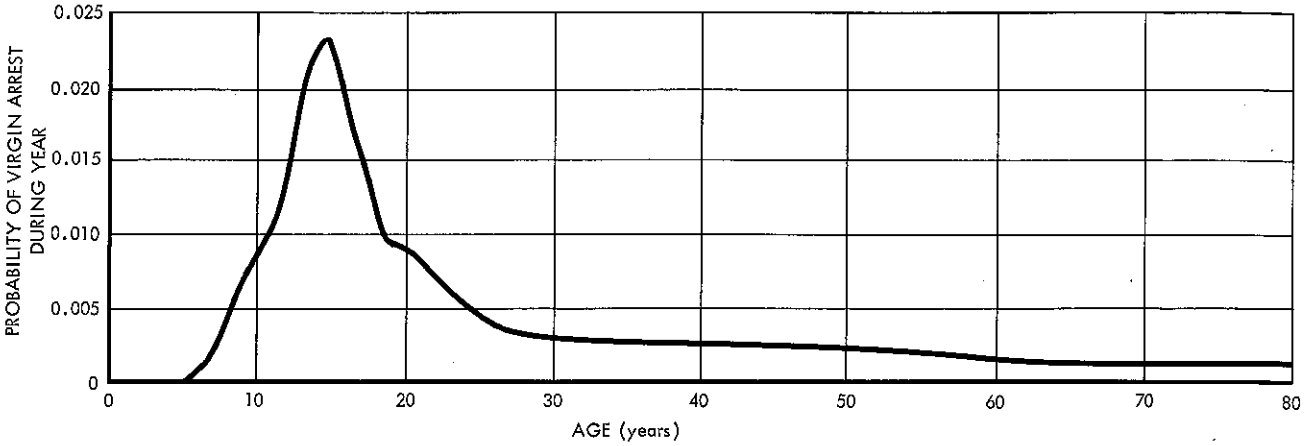
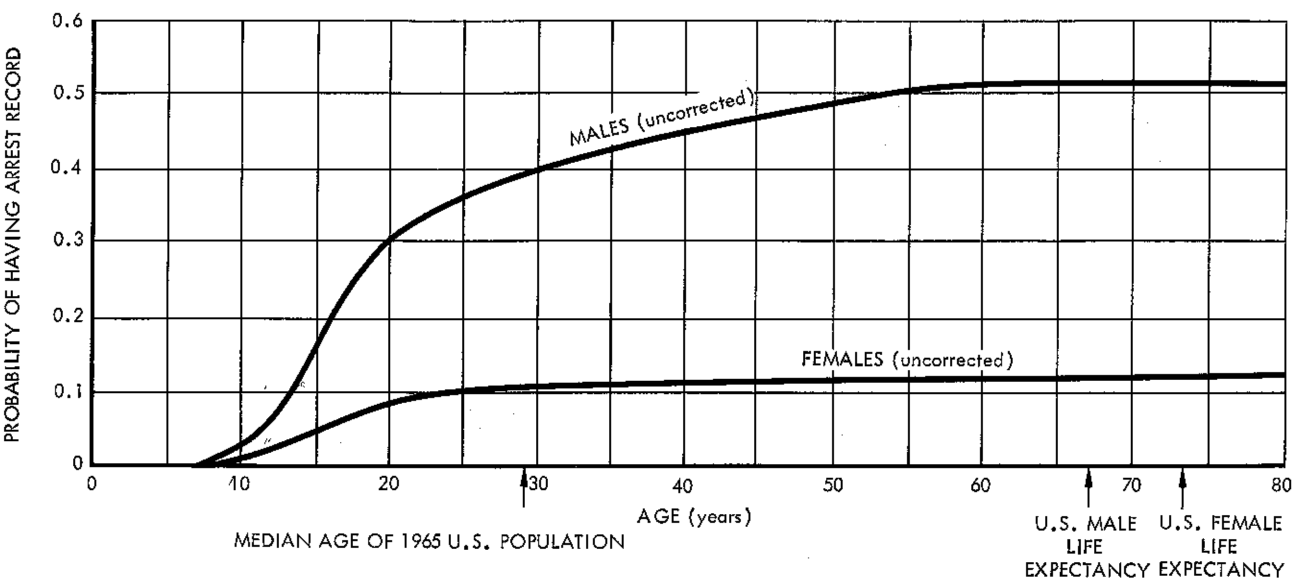


FIGURE J-7. PROBABILITY OF AN INDIVIDUAL 5-YEAR OLD BEING ARRESTED FOR A NONTRAFFIC OFFENSE BY THE TIME HE REACHES GIVEN AGE, ASSUMING FUTURE ARREST PROBABILITIES REMAIN THE SAME AS THOSE IN 1965



rent population median age of 29.5) since the probability of being arrested was lower in the years prior to 1965. A statewide representative sample of 11,329 Minnesota boys and girls (28 percent of all 9th-grade public school children for the 1953–54 school year) showed that by the time they reached 17½ years of age 24.2 percent of the boys and 6.3 percent of the girls had either police or court records for offenses more serious than a minor difficulty with the police, such as a traffic contact.[14] Using 1965 arrest probabilities, figure J–7 shows that, to an age of 17½ years, our analysis gives an estimate of about 25.1 percent for males and 6.5 percent for females.

Statistical research of the County Court of Philadelphia, together with school census figures and U.S. census data, shows that as of 1961 about 21.4 percent of Philadelphia boys and 7.1 percent of the girls were referred to courts

[14] Hathaway, Stein, Elio Monachesi, and Lawrence A. Young, Journal of Criminal Law, Criminology and Police Science, *50*, 433–440 (January–February 1960).

before reaching age 18.[15] In Fayette County, Ky., as of 1960 it has been estimated that 20.7 percent of the boys and 5.2 percent of the girls were referred to juvenile court before age 18.[16] Based on data from a representative nationwide sample of juvenile courts, it has been estimated that about 1 in 6 boys and about 1 in 23 girls in the country will be brought into juvenile court for delinquency before 18 years of age.[17]

To distinguish city, suburban, and rural populations on the curves in figure J–8, the appropriate correction factors are:

	City	Suburban	Rural
Males	1.190	0.616	0.431
Females	1.177	.584	.305

For example, the city male correction factor was calculated as follows:

$$\left(\frac{\text{Number UCR city male arrests in 1965}}{\text{Number UCR male arrests in 1965}}\right) \times \left(\frac{\text{1965 UCR population}}{\text{1965 UCR city population}}\right) \times \left(\frac{\text{U.S. male-fraction}}{\text{Urban male-fraction}}\right)$$

$$=\left(\frac{3,928,314}{4,431,625}\right)\left(\frac{134,095,000}{101,652,000}\right)\left(\frac{0.492}{0.484}\right)$$

$$=1.190$$

As an illustrative application, consider a male who was 10 years old in 1965 and who can be expected to live in the city for the next 20 years. The probability that he will be arrested at least once by the time he is 30 is:

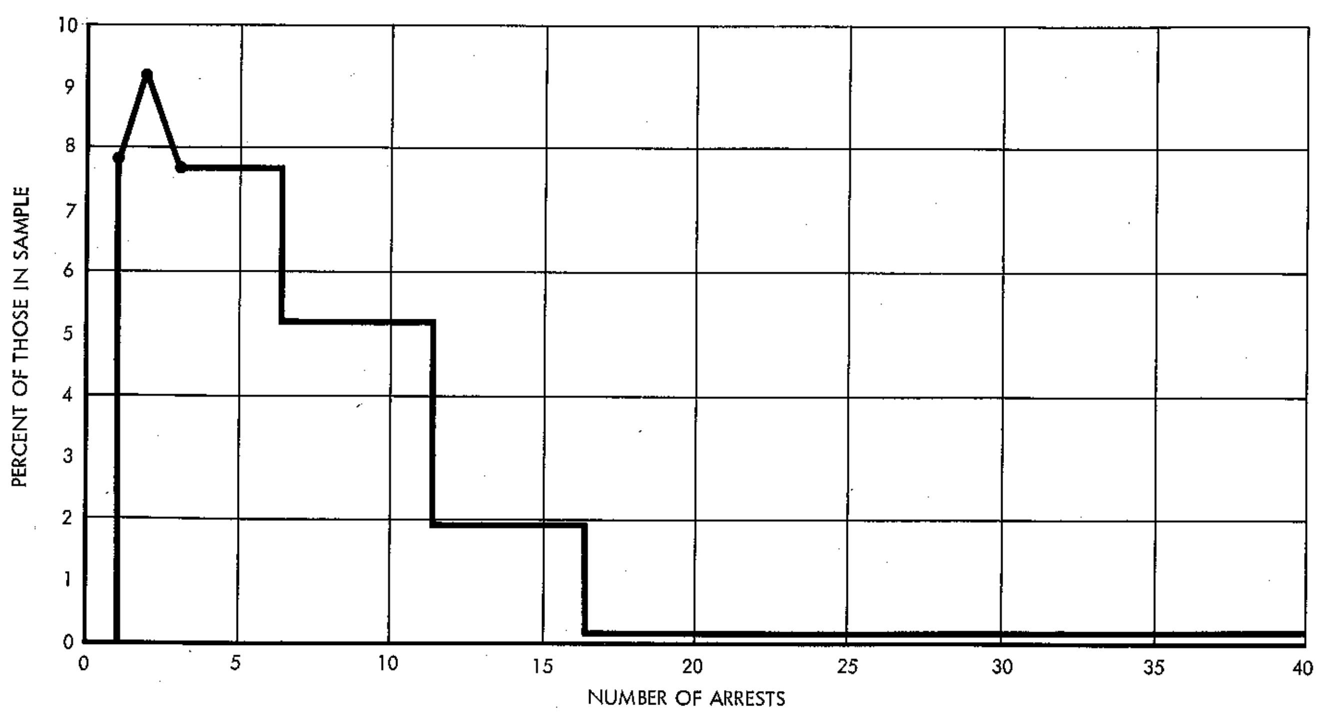
$$1.190\ (0.40+0.03)=0.44$$

Using the above correction factors to provide geographic distinctions for the population, we estimate the lifetime probability of a nontraffic arrest to be:

	Male (percent)	Female (percent)
United States in general	50	12
City	62	15
Suburban	32	7
Rural	22	4

The "United States in general" figures were obtained from the specific figures for "city," "suburban," and "rural" by allocating 67 percent of the U.S. population to city, 12.6 percent to suburban, and 20.4 percent to rural.[18]

FIGURE J-8. ARREST HISTORY FOR SAMPLE OF CONVICTED ADULT FELONS



[15] 48th Annual Report of the County Court of Philadelphia, 1961, p. 134; Monahan, Thomas P., "On the Incidence of Delinquency Social Forces," *39*, 66–72 (October 1960).

[16] Ball, John C., Alan Ross, and Alice Simpson, "Incidence and Estimated Prevalence of Recording Deliquency in a Metropolitan Area," American Sociological Review, *29*, 90–93 (February 1964).

[17] Perlman, I. Richard, "Juvenile Court Statistics," 1964; Children's Bureau Statistical Series, No. 83, HEW, 1965, p. 1.

[18] 1965 UCR, p. 44. Actually, the 67 percent refers to standard metropolitan statistical areas; "suburban," as used in the UCR, includes some counties lying within such areas but excluding the core city. However, the UCR gives insufficient information to improve the allocation.

This then corrects the overall results to roughly account for the disproportionate urban reporting to the UCR.

If we wanted to break these figures down finer, the next logical factor to consider would be economic level. But data correlating arrests to income are not available. However, some racial breakdown data are available, and this does provide limited information on the effects of economic level. Once factors such as sex, income, and residence are considered, any additional differences due to race are small.

To get a racial breakdown, we correct for the number of arrests by race as recorded in the UCR, the population by race as estimated by the Census Bureau, and the basic virgin arrest fraction by race taken as 15 percent for whites and 8 percent for nonwhites. The results are tabulated below:

Males	Percent	Females	Percent
U.S. Males	50	U.S. Females	12
City	62	City	15
White	58	White	14
Nonwhite	[19] >90	Nonwhite	25
Suburban	32	Suburban	7
White	30	White	7
Nonwhite	55	Nonwhite	12
Rural	22	Rural	4
White	21	White	3
Nonwhite	38	Nonwhite	7

[19] The data used was not sufficiently accurate to justify making other than general estimates of percentages exceeding roughly 80 percent. In this case, effects such as differential survival probabilities between males and females and whites and nonwhites, which were not taken into consideration, as well as the basic uncertainties in the virgin arrest ratios can produce distortions with greater absolute magnitudes.

CONVICTIONS

The data problems which limited the previous analysis on arrests was serious, but could be overcome by taking a number of approaches and by trying various data sources. In trying to extend the analysis to include convictions rather than only arrests, the data problems became even more severe, and very crude approximations became necessary. Despite these limitations, that extension is made here, largely to illustrate the approach, to provide a first order-of-magnitude estimate, and to identify the data still needed.

The conviction analysis is based on integration over age of virgin conviction probabilities. Since age-specific conviction probability data is not available, we carry out this integration by summing over the average virgin conviction probabilities for a time-slice of the population. This amounts to assuming that the population-age distribution is flat; i.e., the same number of people falling in each age group, with nobody in age groups past the expected remaining lifetime T beyond the age at which we start our cohort. If the total population is P_{total}, then the number in any age group in this model is $M = P_{\text{total}}/T$. The lifetime conviction probability is then:

$$P = \frac{r_c C}{M} = \frac{r_c C}{P_{\text{total}}/T}$$

where C is the number of convictions per year and r_c is the fraction of convictions which are of persons never before convicted.

The actual population-age distribution is more dense at the young ages. Thus, since most virgin convictions occur toward the early part of the life expectancy, the flat distribution assumption, which was necessary because of the unavailability of national age-specific conviction data, will result in an overestimate of lifetime conviction probabilities. To estimate the amount of error introduced by the flat distribution assumption, this assumption can be applied to the arrest calculations. This would have resulted in estimates of 64 and 17.5 percent for lifetime arrest probabilities of males and females, respectively, instead of 52 and 13 percent which resulted from the actual distribution. Since most virgin convictions occur within a few years after most virgin arrests, it is reasonable to estimate the actual lifetime conviction results by introducing correction factors of $\frac{52}{64} = 0.81$ for males and $\frac{13}{17.5} = 0.74$ for females.

The probability of an adult being charged with a part I offense[20] and being convicted (of the charged offense or a lesser offense) in an adult court in 1965 was:[21]

$$\left(\frac{\text{1965 part I city convictions}}{\text{1965 UCR city population}}\right) \times \left[\frac{\dfrac{\text{city population}}{\text{total population}}}{\dfrac{\text{city arrests}}{\text{total arrests}}}\right] \times \left(\frac{\text{1965 U.S. population}}{\text{1965 U.S. population} \geq 18}\right)$$

$$= \left(\frac{120{,}736}{56{,}554{,}000}\right) \left[\frac{\dfrac{101{,}652{,}000}{134{,}095{,}000}}{\dfrac{4{,}401{,}598}{4{,}955{,}047}}\right] \left(\frac{194{,}400{,}000}{124{,}099{,}000}\right)$$

$$= 0.00285$$

Next we need the fraction of those convicted who do not have a prior conviction record. For a sample of 88 U.S. District Courts in 1964 this was 0.349, being 0.326 for males and 0.609 for females. These numbers would probably be unrepresentative of many UCR-type crimes (e.g., drunkenness which has a very high recidivism rate). Lacking better data, however, we will make our preliminary calculations using these numbers, recognizing that the final results may be in error.

Further, we assume for purposes of making a first estimate that male and female conviction probabilities are the same. Then, assuming conviction probabilities do not change in future years, the probability of an indi-

[20] These offenses are criminal homicide (murder and nonnegligent manslaughter, and manslaughter by negligence), forcible rape, robbery, aggravated assault, burglary—breaking or entering, larceny—theft, and auto theft.

[21] Different tables in the UCR have different population bases. Hence, it has been necessary to include appropriate population terms, since individual ratios must use the population base of the particular table.

vidual who was 18 in 1965 eventually being charged with a part I offense and convicted is:

Males:

$$(0.00285)\begin{pmatrix}\text{Fraction of 1965 part I}\\ \text{offense arrests which}\\ \text{were of males}\end{pmatrix}$$

$$\left(\frac{\text{Male first conviction fraction}}{\text{Male fraction of population}}\right)$$

$$\times(\text{remaining male life expectancy})$$

$$=(0.00285)(0.866)\left(\frac{0.326}{0.492}\right)(67-17)$$

$$=0.0818$$

Females:

$$(0.00285)\begin{pmatrix}\text{Fraction of 1965 part I}\\ \text{offense arrests which}\\ \text{were of females}\end{pmatrix}$$

$$\left(\frac{\text{Female first conviction fraction}}{\text{Female fraction of population}}\right)$$

$$\times(\text{remaining female life expectancy})$$

$$=(0.00285)(0.134)\left(\frac{0.609}{0.508}\right)(73-17)$$

$$=0.0256$$

Now, applying the sex-residence correction factors given previously, we can calculate the lifetime probability, by sex and residence, of an adult conviction, which results from being charged with a part I offense:[22]

	Males, percent	Females, percent
City	7.9	2.4
Suburban	4.0	1.2
Rural	2.8	0.6

Table J–2.—Calculation of Convictions per Arrest

	All arrests (percent)	Part I arrests ÷ all arrests	Part I 1st arrest fraction ÷ all 1st-arrest fraction	Part I arrests (percent)	Convictions on a part I charge (percent)	Convictions per part I arrest
Males:						
City	62	0.158	2	19.6	7.9	0.40
Suburban	32	.201	2	12.9	4.0	0.32
Rural	22	.193	2	8.5	2.8	0.33
Females:						
City	15	.189	2	5.7	2.4	0.39
Suburban	7	.219	2	3.1	1.2	0.36
Rural	4	.130	2	1.0	0.6	0.52

We make the assumption that the part I first-arrest fraction divided by the fraction of all arrests that are first arrests is approximately two. This is based on the data that one-fourth of the 1965 entries in the FBI criminal career file were new entries and one-eighth of the total U.S. arrest in 1965 were first arrests. We now estimate the fraction of convictions per part I arrests.[22] The calculation is tabulated in table J–2.

Next, we assume that the fractions of convictions per arrest for part I arrests are roughly typical of the fractions for all arrests.[23] Then, using these estimates of the fractions of convictions per arrest, we can make the following order-of-magnitude estimates of the lifetime conviction probabilities for all UCR-type charges:

	Arrests (percent)	Convictions per arrest	Corrected lifetime convictions percentage
Males			19
City	62	0.40	25
Suburban	32	.32	10
Rural	22	.33	7
Females			5
City	15	.39	6
Suburban	7	.36	3
Rural	4	.52	2

The reasonableness of these results, which imply that a fraction of $f\approx0.12$ of the population is eventually convicted, can be seen by the following consideration. Let P be the total U.S. population, g be the fraction which is at any point in time under supervision of some sort in the correctional system, and T be the total number of years under correctional supervision during the life of an average individual in the fraction f. Then, since the average lifetime is about 70 years:

$$\frac{70}{T}gP=fP$$

From correctional[24] data, $g\approx0.007$.

Therefore:

$$f\approx\frac{0.49}{T}$$

Thus, $f\approx0.12$ implies that $T\approx4$ years. Although insufficient data has been found to permit an independent calculation of T, this does appear to be in the realm of the reasonable.

LIFETIME ARREST HISTORY PROFILES

We will now estimate lifetime arrest history profiles for the U.S. population. First, let us look at the male population.

One estimate of the profile can be obtained by assuming that after the first arrest, subsequent arrests are statistically independent in the sense that their frequency of

[22] Corrected for age distribution by including a factor 0.81 for males and 0.74 for females, as previously discussed.

[23] Since part I charges represent the more serious charges, this is probably a conservative assumption.

[24] See chapter 4. This includes correctional detention, probation, institutionalization of all kinds, and parole and other aftercare.

occurrence is given by the Poisson distribution. Then if p is the probability of being arrested at least once, the probability of being arrested n times is given by pq_{n-1}, where:

$$q_n=\frac{S^n}{n!}e^{-S}$$

is the probability of n subsequent arrests given a first arrest, and:

$$S=\frac{1}{r}-1$$

is the average number of subsequent arrests of those arrested at least once.

For $p=0.5$, $r=0.125$, the resulting curve is plotted on figure J–9 (dotted line), where the vertical axis is number of arrests (n) and the horizontal axis is arrest-nonproneness (x) measured in percentile of the population arranged in order of decreasing number of arrests. For example, an individual having arrest-nonproneness $x=30$ percent means that 30 percent of the male population is arrested at least as many times as he is.

Another estimate of the profile can be obtained from District of Columbia Crime Commission data. The percent of the sample of convicted adult felons with a history of a specific number of arrests versus the number of arrests [25] is plotted on figure J–8. Subdividing the $p=50$ percent accordingly, we obtain the dashed curve on figure J–9. Although convicted adult felons may not represent the arrested population generally in terms of number of arrests, this appears to be the best data available. The points on the dashed curve, drawn in step fashion, can be approximated rather closely by a simple exponential function of the form

$$x=50e^{-n/a}$$

where a is the mean number of arrests of those who have been arrested at least once. The fact that a is the mean can be verified in the equation:

$$50a=\int_0^\infty x\,dn$$

Using 1965 UCR data, the area under the curve should be:

$$A=\left(\frac{\text{males arrested per year}}{\text{male population}}\right)(\text{average male lifetime})$$

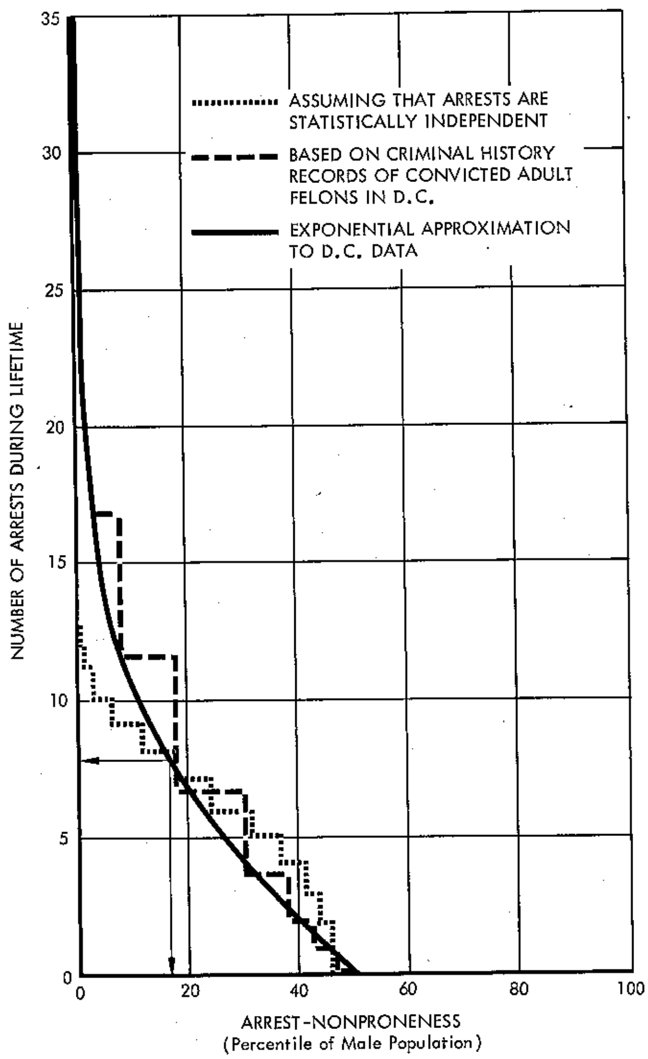
$\times$ age distribution correction factor/geographic correction factor

$\times$ 100 percent

$$=\frac{(4{,}431{,}625)(67)}{(0.492)(134{,}095{,}000)}\times\frac{0.81}{0.96}\times 100$$

$$=382$$

FIGURE J-9. PROJECTED ARREST HISTORY PROFILE FOR U.S. MALE POPULATION



Using this datum, we can determine a.

$$\int_0^\infty x\,dn=50a=382$$

$$\therefore a\approx 7.6$$

This is consistent with our previous estimate that about one-eighth of all arrests are first arrests.

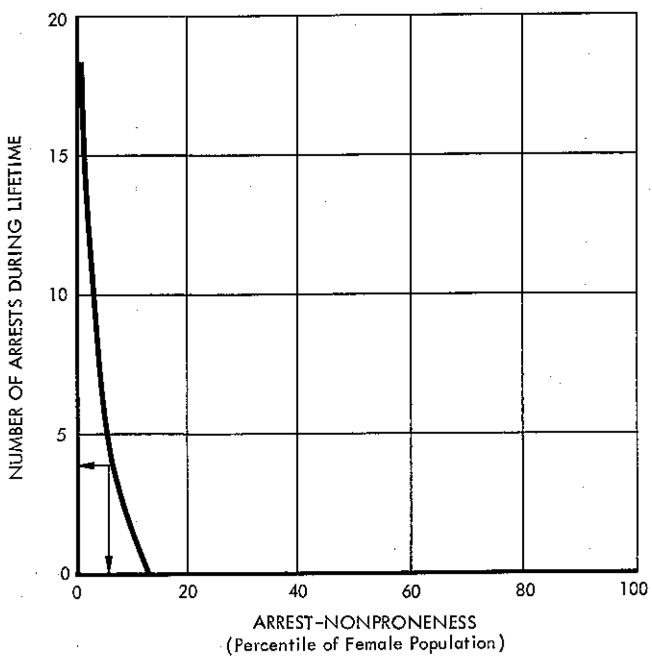
The curve:

$$x=50e^{-n/7.6}$$

is the solid line on figure J–9.

[25] Report to the President on Crime in the District of Columbia, 1966, app., p. 591.

FIGURE J-10. PROJECTED ARREST HISTORY PROFILE FOR U.S. FEMALE POPULATION



Suppose, for example, we define strong repeaters as those who experience more than the average number of arrests (7.6) for nontraffic offenses in their lifetimes. The percentage of the male population who are strong repeaters is then:

$$x_s = 50e^{-1}$$

$$\approx 18 \text{ percent}$$

From figure J–9, we can also calculate the probabilities p_i, given the individual is arrested at least i times, that he is arrested at least $i+1$ times.

We have already established that $p_0 \approx 0.5$. To find p_i for $i \geqq 1$, we set $i = n+1$ and evaluate:

$$p_i = \frac{x(n+1)}{x(n)} \qquad \text{(by the definition of } x\text{)}$$

$$= e^{-1/a}$$

Using $a=7.6$:

$$p_i \approx 0.88 \text{ for all } i \geq 1$$

These results show that there is a striking difference between those people who get arrested at least once and those who do not. The strong correlation between the first arrest and subsequent arrests, which shows up in the fact that $p_i \gg p_o$ for $i \geqq 1$, indicates that there is a phenomenon of arrest-proneness. Were there no correlation between arrests, then they would be Poisson distributed over the whole population and not just the arrested population. If so, then we would expect the average number of arrests of those arrested at least once to be only:

$$\frac{-ln(1-p_o)}{p_o} = \frac{-ln(1-0.5)}{0.5}$$

$$\approx 1.36$$

In reality it is about eight, a much larger number. So arrests are not Poisson distributed over the whole population. In fact, there is even a correlation between subsequent arrests, since the distribution over the arrested population is closer to an exponential distribution than to a Poisson distribution, as we have seen.

Repeating this same calculation for females, we have:

$$A' = \frac{(599{,}768)(73)}{(0.508)(134{,}095{,}000)} \times \frac{0.74}{0.96} \times 100$$

$$= 49.8$$

$$a' = \frac{49.8}{13.0}$$

$$\approx 3.8$$

If arrests were uncorrelated, one would expect a' to be

$$-\frac{ln(1-0.13)}{0.13} \approx 1.96$$

The resulting curve:

$$x' = 13e^{-n'/3.8}$$

is plotted on figure J–10. Using the same definition, the percentage of strong repeaters among the female population is:

$$x_s' = 5 \text{ percent}$$

LIFETIME CONVICTION HISTORY PROFILES

In the previous section, we estimated lifetime arrest history profiles. We now turn to convictions. For males, the relevant data is:[26]

$$C = \text{(male convictions per male arrest) } A$$

$$= (0.36)382$$

$$= 137$$

$$a'' = \frac{137}{19}$$

$$= 7.2$$

[26] This says that the average number of convictions during the life of those convicted at least once (including convictions for such things as drunkenness), assuming 1965 conviction rates hold into the future, is expected to be 7.2 for males.

The resulting curve

$$y = 19e^{-N/7.2}$$

is the upper curve on figure J–11. The corresponding curve

$$y' = 5e^{-N'/4.1}$$

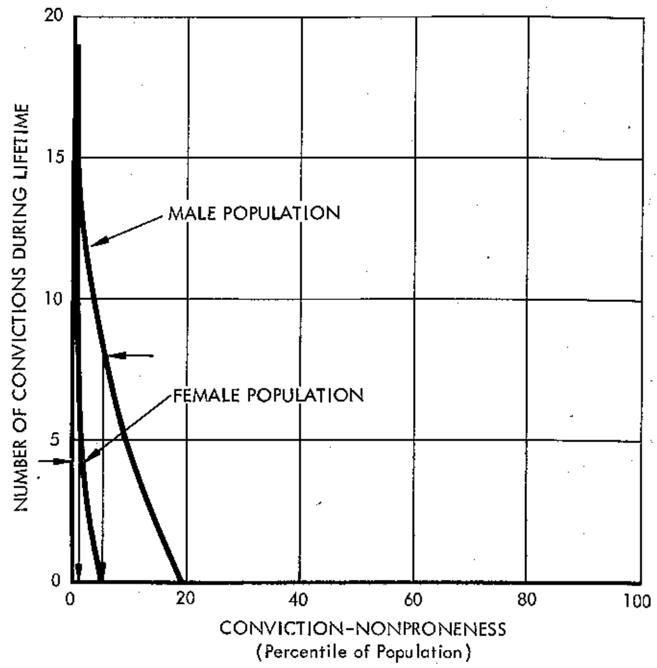
for females is the lower curve on Figure J–11.

Using a definition of "strong repeater" for convictions similar to that for arrests, we obtain

$$y_s \approx 7 \text{ percent}$$
$$y'_s \approx 2 \text{ percent}$$

as the percentage of the male and female populations, respectively, who are strong repeaters in convictions (over 7.2 lifetime convictions for males and 4.1 for females).

FIGURE J-11. PROJECTED CONVICTION HISTORY PROFILE FOR U.S. POPULATION



U. S. GOVERNMENT PRINTING OFFICE : 1972 O - 464-220