Estimating the Extent of Illicit Drug Abuse in New Jersey Using Capture-recapture Analysis

Submitted by:
Division of Addiction Services
New Jersey Department of Human Services
July 2005

DAS – DHS

Acknowledgements

This report is an update of the publication which was originally published by Abate Mammo, Ph.D., in November 1998. This same model was applied on current data to examine recent trends in the estimated number of drug abusers in New Jersey.

Reproduced by:

Office of Research, Planning, Evaluation and Information Systems
Division of Addiction Services
NJ Department of Human Services

Introduction

The ideal approach to treatment need estimation is the use of an objective quantitative assessment that allows for the diagnosis of the level of substance abuse or dependence. Such a diagnostic scheme (e.g., DSM-IV), however, is limited by the ease of access to the study population and the will of this population to participate in often probing interviews that could easily be perceived to have legal consequences as well as stigma. Unlike for alcohol and cigarette use or abuse, such approaches usually fail to provide adequate data on the use or abuse of illicit drugs by householders.

The search for quantitative data that could be used for some comprehensive treatment needs assessment for the State and for its counties led us to apply the two sample capture-recapture model. The model output shows that a substantial number of people abuse illicit drugs in the State and that the county distribution of abusers is consistent with our expectations.

Background

A method originally used for the study of salmon populations by Walton (1653) and formally introduced for the estimation of the size of populations using multiple independent samples taken over time by Peterson (1894) has now been qualified to provide estimates of similar populations that do not lend themselves to standard methods of estimation techniques (see Cormack (1968) for an extensive review of the statistics of this method). Later, the independence assumption was relaxed by providing methods of estimating dependencies (Bishop, Fienberg and Holland, 1975; Doscher and Woodward, 1983). This study concerns the estimation of a closed population (i.e., change in the population occurs because of births or deaths) and follow the notation of Bishop et al. (1975). Models that entertain open populations are reviewed extensively elsewhere (Cormack, 1968).

The capture-recapture method assumes that there is a set of samples, lists, or rosters containing subjects that are uniquely identified so that one can determine whether an individual was present or absent in any one of the samples. If there are k sources of data (i.e., samples) collected at non overlapping time points and not necessarily taken in any sequential order, it is assumed that the individual will fall into any one of the 2^k-1 cells and that the 2^kth cell is not observed in any of the k samples. The statistical problem is to estimate the size of the population which is known to exist but

has not been possible to sample.

To familiarize the reader with the notations in such a model, one can assume that there were two samples taken and the following cross-classifications were obtained.

First Sample (2002)		Second Sample (2004)		
		Admitted	Not Admitted	Total
	Admitted	x_{11}	x_{12}	x_{1+}
	Not admitted	x_{21}	-	
	Total	x_{+1}		

x_{ij} corresponds to the observed number of individuals in cell (i,j). The subscript 1 indicates that the individual is observed (admitted) in the sample and 2 indicates that it was not observed (not admitted) in the sample. For example, x_{21} represents the number of individuals who were observed in the second sample (i.e., 2004) but not in sample one (i.e., 2002). Notice that x_{22} is the missing count in the cell designated by " -". Our objective is to estimate this missing count of individuals who were not reported in any of the two samples.

If we let n be the total number of individuals observed in the 2^k-1 cells (3 cells, in the example), then

$$\mathbf{n = x_{11} + x_{12} + x_{21} = \sum \sum * x_{ij}} \qquad \mathbf{(1)}$$

The asterisk indicates that the observation in cell (2,2) is not included.

Let p_{11} be the probability of an individual being admitted in both 2002 and 2004, p_{1+} the probability of being admitted in 2002, p_{+1} the probability of being in 2004. If we assume that the two samples (or admissions in 2002 and 2004) are independent then $p_{11} = p_{1+}\, p_{+1}$, and if n is fixed then (x_{11}, x_{12}, x_{21}) has the multinomial distribution with probability function given in equation 2.

$$\begin{pmatrix} n \\ x_{11}, x_{12}, x_{21} \end{pmatrix} \frac{(p_{1+}\, p_{+1})^{x11}[p_{1+}(1-p_{+1})]^{x12}[p_{+1}(1-p_{1+})]^{x21}}{[1-(1-p_{1+})(1-p_{+1})]^{n}} \quad \textbf{(2)}$$

The maximum likelihood estimates of the marginal probabilities, p_{+1} and p_{1+} are given in Equation 3.

$$p_{1+} = x_{11}/x_{+1} \qquad p_{+1} = x_{11}/x_{1+} \quad \textbf{(3)}$$

Suppose that n has a binomial distribution with sample size N and the probability of being observed in at least one of the two samples, $p = 1-(1-p_{1+})(1-p_{+1})$. Then the probability of selecting n out of N is

$$\begin{pmatrix} N \\ n \end{pmatrix} (p)^{n}(1-p)^{N-n} \quad \textbf{(4)}$$

The maximum likelihood estimate (MLE) of N ($\hat{N}$) is then given as the number of observed individuals, n divided by p. If we then substitute x_{11}/x_{+1} for $\hat{p}_{1+}$ and x_{11}/x_{1+} for $\hat{p}_{+1}$ into the MLE equation we get

$$N = (x_{1+}x_{+1})/x_{11} \quad \textbf{(5)}$$

Peterson (1894) derived the same formula to estimate the size of fish populations. Sekar and Deming (1949) considered situations where x_{+1} and x_{1+} were not fixed and showed that combining the multinomial probability function (Equation (2)) with the binomial probability function (equation (4)) will result in the same MLE estimators for N, p_{1+} and p_{+1}. They also provide a formula (equation (6)) for the asymptotic variance of N (Bishop et al., 1975).

$$\hat{Var}(\hat{N}) = \frac{x_{12}\, x_{21}\, x_{1+}\, x_{+1}}{x_{11}^{3}} \quad \textbf{(6)}$$

In some situations, the presence or absence of an individual in the second sample may depend on whether or not that individual was present in the first sample. Naturally, this violates the independence assumption used earlier between the two samples. Models are available that correct

biases that may originate from the lack of independence of samples[1].

For k samples, where there are 2^k incomplete cross-classifications with one missing cell, the estimation formulas can be generalized. Let $m(i)_{12..k}$ be the expected number of individuals in the $(i_1,i_2,...,i_k)$ cell of the 2^k table, where i_j (j=1,2,...,k) equals 1 if the individual is present in this cell or 2 if absent. Cell (2,2,...,2) is the missing cell so that $m_{22...2} = 0$. Suppose also that the set S contains the 2^k-1 cells excluding the cell (2,2,...,2). Then n equals to $\Sigma^* x(i)_{12...k}$ where the summation runs over the set S. The probability that an individual falls in the cell $(i_1,i_2,...,i_k)$ is $m(i)_{12...k}/n$. If N is the total number of individuals in the population, N - n individuals are absent from the k samples. Following the 2-sample example we will proceed to estimate $m_{22...2}$ by $m^*_{22...2}$ as follows.

$$m^*_{22...2} = \frac{M_{odd}}{M_{even}} \qquad \textbf{(7)}$$

In equation (7), M_{odd} is the product of all $x(i)_{12...k}$ in S where the sum of the subscripts is equal to an odd number. M_{even} likewise refers to the product of those with even sums in their subscripts. The estimation of N and its asymptotic variance follow similar expressions as in the two sample model. The estimate for N is then given as follows:

$$N = n + m^*_{22...2} \qquad \textbf{(8)}$$

The capture-recapture model was applied to our Client Oriented Data Acquisition Process (CODAP) data by French (1977a, 1977b) to estimate the size of heroin abusers in New Jersey. Bonett, Woodward and Bentler (1986) recently suggested a linear model for the estimation of the size of a closed population using multiple recapture samples. Doscher and Woodward (1983) caution, however, that an attempt to estimate the size of a heterogenous population (different sampling probabilities) without stratification of the sample into homogenous groups would bias

[1] The log-linear model which assumes that all pair-wise relationships are present is given as

$$\log m_{ij} = u + u_1(i) + u_2(j) + u_{12}(ij)$$

Where u = the grand mean of the logits of the expected cell counts
$u_1(I)$ = main effect of variable 1
$u_2(j)$ = main effect of variable 2
$u_{12}(ij)$ = two factor effect between variables 1 and 2

estimates arrived at using this method and suggest methods of correction for such problems. Following the advice by Doscher and Woodward (1983), Mammo (1995) estimated the numbers of heroin, cocaine and other drug abusers in New Jersey and for its 21 counties. This report closely follows his approach from 1995 with minor exceptions.

Assumptions Used

We applied the two sample Capture-recapture method to data obtained from the 2002 and 2004 combined NJSAMS and ADADS[2] data sets to estimate the number of drug abusers in New Jersey. The following assumptions were made for groupings of the treatment population in order to maximize independence:

1) An illicit drug abuser is one who was admitted for at least one illicit drug problem as primary, secondary or tertiary drug of choice. We excluded alcohol only admissions from the analysis.
2) Admissions for a particular drug abuse treatment in a county in 2002 and 2004 were independent of each other (i.e., the fact that drug abusers were admitted in 2002 had no influence on their readmission probabilities in 2004).
3) Substance abusers who seek treatment have characteristics similar to those who do not.
4) A heroin abuser is one who was admitted to treatment for a heroin or opiate abuse problem as the primary, secondary or tertiary drug of choice.
5) A cocaine abuser is one who was admitted to treatment for a cocaine or crack problem as the primary, secondary or tertiary drug of choice after excluding heroin abusers and alcohol only abusers.
6) Other drug abusers are those who were admitted for treatment for drugs other than heroin, cocaine and crack.

The above assumptions imply that treatment admissions for a particular drug are independent of each other, and abusers who sought treatment had characteristics similar to those who did not.

2 ADADS was the surveillance system maintained by the New Jersey Division of Addiction Services (before 2003) to monitor treatment activities for substance abuse and dependence in the State. It was replaced by NJSAMS, a web-based data reporting system for substance abuse treatment providers in New Jersey.

Unduplicated admissions data from 2002 and 2004 were used to avoid overlaps between the two samples. Almost all the 2002 admissions were discharged from treatment before 2004 making them available for readmission (recapture) in 2004. To improve homogeneity as advised by Doscher & Woodward (1983), separate estimates were made for heroin, cocaine, and other drugs within each county. This allows us to assume equal probability of capture in the two samples for each drug type within a county thereby decreasing heterogeneity of populations.

Findings

Applying the capture-recapture method to treatment data provides useful estimates for the number of illicit drug abusers, the majority of whom may end up in treatment[3]. The validity of this method for the estimation of other substances such as alcohol is questionable. Household surveys are believed to produce more reliable estimates for alcohol abuse than for illicit drugs because of social desirability.

We estimate that in 2003, there were 86,495 (95% C.I.: 84,588 to 86,647) heroin abusers in need of treatment in New Jersey up from 86,353 estimated in 1998. As expected, Essex County contributes the largest statewide estimated number of heroin abusers with 19,874 followed by Camden (7,405) and Hudson (7,164) Counties. The respective estimates in 1998 were 22,750; 6,530 and 8,703. Table 1 presents details of the 2003 estimates for each county for heroin, cocaine, and other drugs estimates.

Unlike the 1998 estimate, the number of cocaine abusers appears to have gone down from the 1998 estimate of 93,741 to 65,959 (95% C. I.: 62,350 to 69,568). Again, Essex County contributed the largest number of cocaine abusers with 6,940 (95% C. I.: 5,317 to 8,563) followed by Burlington (6,720), Hudson (5,320) and Monmouth (5,082) Counties. Consistent with 1998 estimates, variations in heroin abuse by county remain higher than variations than cocaine abuse. See Tables 1 to 3 and Charts 2.21 to 2.22.

[3] We expect most drug abusers to end up in treatment through personal choice, through contact with the criminal justice system or through other factors.

Chart 2.21 Percent of the Adult Population in Need of Treatment for Illicit Drug use in New Jersey

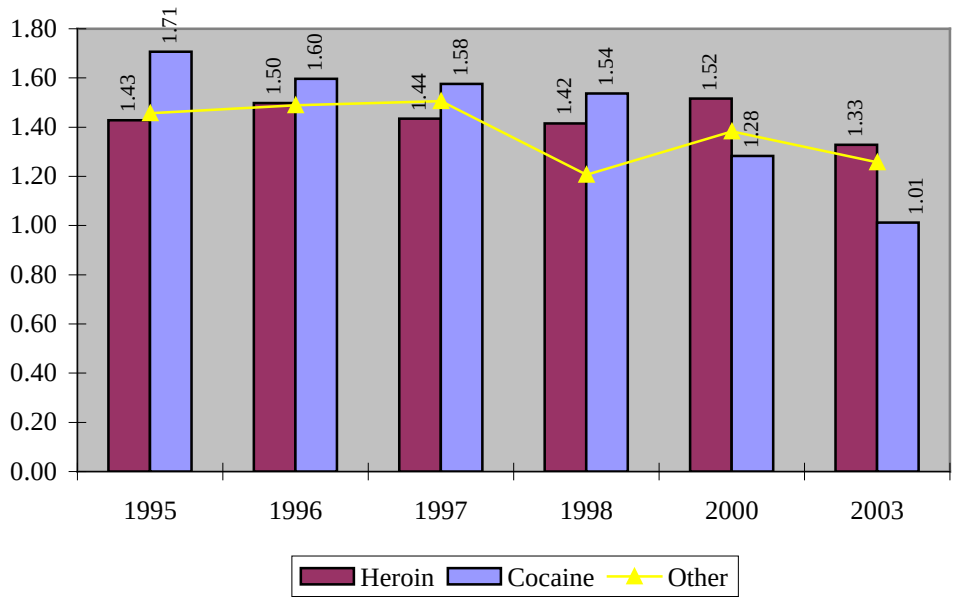
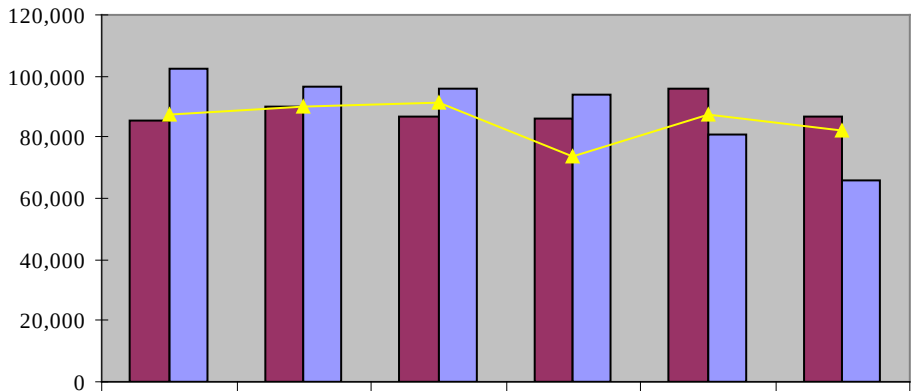


Chart 2.22 Number in Need of Treatment for Illicit Drug Use in N. J.



	1995	1996	1967	1998	2000	2003
Heroin	85,640	90,251	87,045	86,353	95,937	86,495
Cocaine	102,359	96,213	95,553	93,741	81,149	65,959
Other	87,303	89,754	91,339	73,635	87,478	81,853

All other drugs combined contribute 81,853 (95% C. I. (75,122, 88,584)) of the 234,307 estimated illicit drug abusers in the state. We believe that the actual prevalence of illicit drug abuse in the state is much larger than the estimates suggest, that may be due to poly drug abuse.

In Table 2 we present the number of injection drug abusers estimated using the two-sample Capture-recapture model. We estimated that there were 28,301 injectors (95% C.I.: 27,501, 29,101) in the state in 2003, almost all of whom were heroin injectors. As expected, Essex County had the largest number of injectors with 4,062 (95% C. I.: 3,747, 4,377) followed by Camden County with 2,444 (95% C. I.: (2,148, 2,740)) and Passaic County with 2,121 (95% C. I.: (1,899, 2,343)).

Table 3 presents the 1998 and 2003 heroin, cocaine, and other drug estimates side by side for comparison purposes. The table shows a growing need for Heroin and other drugs treatment while the need for cocaine treatment has declined.

Discussion

Our search for quantitative information that could be used for a more comprehensive treatment need assessment for the state and for its counties prompted us to apply the two sample capture-recapture model. We find that a substantial number of people (234,307) are in need of treatment for their illicit drug abuse problems in the state. Of these, 86,495 need treatment for their heroin problems, 65,959 need treatment for their cocaine problems and 81,853 need treatment for other drug problems. Unlike the 1998 estimates, there are fewer cocaine abusers in the State than heroin abusers in 2003. Treatment of other drug abuse appears to have gone up from 73,635 to about 81,853 during the same period.

Consistent with expectation, there is a substantial variation in heroin abuse by county with Essex County leading all counties in heroin, cocaine abuse, and other drug abuse.

The estimates presented here will supplement other studies in the Treatment Needs Assessment family of studies such as the 2003 Telephone Household Survey in assessing need and demand for treatment in the State and its counties.

Table 1

Number of people who abuse illicit drugs in New Jersey, 2003

	Heroin			Cocaine			Other Drugs			Total
	Number in Need	95% Confidence Lower	95% Confidence Upper	Number in Need	95% Confidence Limits Lower	95% Confidence Limits Upper	Number in Need	95% Confidence Lower	95% Confidence Upper	
Atlantic	3,960	3,554	4,366	2,441	1,752	3,130	6,108	1,992	10,224	12,509
Bergen	3,844	3,406	4,282	4,418	3,620	5,216	4,719	3,772	5,666	12,981
Burlington	2,379	1,897	2,861	6,720	1,910	11,530	2,982	(1,230)	7,194	12,081
Camden	7,405	6,622	8,188	4,428	3,550	5,306	6,973	5,809	8,137	18,806
Cape May	1,305	1,018	1,592	1,344	854	1,834	2,612	958	4,266	5,261
Cumberland	1,357	1,074	1,640	2,467	1,800	3,134	2,233	1,155	3,311	6,057
Essex	19,874	18,996	20,752	6,940	5,317	8,563	8,218	6,533	9,903	35,032
Gloucester	2,310	1,967	2,653	2,043	1,361	2,725	3,700	1,580	5,820	8,053
Hudson	7,164	6,603	7,725	5,320	3,943	6,697	4,742	3,916	5,568	17,226
Hunterdon	753	546	960	479	239	719	1,034	205	1,863	2,266
Mercer	2,636	2,251	3,021	3,427	2,619	4,235	4,347	2,029	6,665	10,410
Middlesex	5,961	5,397	6,525	4,361	3,122	5,600	5,627	3,619	7,635	15,949
Monmouth	4,831	4,456	5,206	5,082	4,316	5,848	7,137	6,311	7,963	17,050
Morris	2,249	1,983	2,515	1,849	1,178	2,520	1,645	572	2,718	5,743
Ocean	4,181	3,857	4,505	2,699	2,046	3,352	4,237	1,914	6,560	11,117
Passaic	5,669	5,212	6,126	3,654	2,716	4,592	4,690	3,990	5,390	14,013
Salem	570	384	756	707	389	1,025	1,239	205	2,273	2,516
Somerset	1,291	1,070	1,512	1,570	720	2,420	1,820	1,376	2,264	4,681
Sussex	1,526	1,162	1,890	1,155	90	2,220	913	(1,188)	3,014	3,594
Union	6,436	5,941	6,931	4,141	3,163	5,119	4,797	3,640	5,954	15,374
Warren	794	642	946	714	445	983	2,080	730	3,430	3,588
New Jersey	86,495	84,588	86,647	65,959	62,350	69,568	81,853	75,122	88,584	234,307

Source: Drug need estimates are made using data from the Alcohol and Drug Abuse Data System (2002 and 2004 ADADS & NJSAMS data) and applying a two-sample capture-recapture estimation approach.

Note: Heroin estimates refer to any heroin admission as primary, secondary or tertiary drug of choice at the time of treatment. Cocaine estimates refer to any cocaine admissions as primary, secondary or tertiary drug of choice at the time of admission. Other drug estimates refer to the residual of drug abusers/dependents after heroin, cocaine and alcohol only are excluded. Both heroin and cocaine estimates are made after alcohol only admissions are excluded.

Table 2

Estimated number of injection drug abusers in New Jersey, 1998 & 2003

	Number of Injectors			
	1998	2003	95% Confidence Interval	
			lower	Upper
Atlantic	1,914	2,044	1,824	2,264
Bergen	1,726	1,425	1,229	1,621
Burlington	695	829	641	1,017
Camden	2,442	2,444	2,148	2,740
Cape May	577	507	388	626
Cumberland	536	691	507	875
Essex	4,336	4,062	3,747	4,377
Gloucester	701	901	746	1,056
Hudson	2,407	2,012	1,813	2,211
Hunterdon	244	266	159	373
Mercer	1,293	1,378	1,161	1,595
Middlesex	1,967	1,944	1,736	2,152
Monmouth	1,543	1,845	1,672	2,018
Morris	610	718	627	809
Ocean	1,237	1,996	1,827	2,165
Passaic	2,022	2,121	1,899	2,343
Salem	100	240	122	358
Somerset	381	380	308	452
Sussex	200	566	403	729
Union	1,770	1,565	1,391	1,739
Warren	275	367	292	442
New Jersey	**26,975**	**28,301**	**27,501**	**29,101**

Table 3

Number of people who abuse illicit drugs in New Jersey, 1998, 2003

County	1998				2003			
	Heroin	Cocaine	Other Drugs	Total	Heroin	Cocaine	Other Drugs	Total
Atlantic	3,624	4,792	5,070	13,486	3,960	2,441	6,108	12,509
Bergen	4,511	5,479	4,426	14,416	3,844	4,418	4,719	12,981
Burlington	1,846	4,420	4,437	10,703	2,379	6,720	2,982	12,081
Camden	6,530	8,109	5,225	19,864	7,405	4,428	6,973	18,806
Cape May	830	1,174	1,895	3,899	1,305	1,344	2,612	5,261
Cumberland	1,012	2,506	1,554	5,072	1,357	2,467	2,233	6,057
Essex	22,750	10,427	6,968	40,145	19,874	6,940	8,218	35,032
Gloucester	1,648	3,050	2,204	6,902	2,310	2,043	3,700	8,053
Hudson	8,703	7,749	5,662	22,114	7,164	5,320	4,742	17,226
Hunterdon	656	641	1,362	2,659	753	479	1,034	2,266
Mercer	2,204	7,081	3,878	13,163	2,636	3,427	4,347	10,410
Middlesex	6,289	6,181	4,231	16,701	5,961	4,361	5,627	15,949
Monmouth	3,796	9,174	7,399	20,369	4,831	5,082	7,137	17,050
Morris	2,656	2,282	1,799	6,737	2,249	1,849	1,645	5,743
Ocean	2,955	3,746	4,130	10,831	4,181	2,699	4,237	11,117
Passaic	6,102	6,182	4,593	16,877	5,669	3,654	4,690	14,013
Salem	186	1,316	663	2,165	570	707	1,239	2,516
Somerset	1,356	1,961	1,873	5,190	1,291	1,570	1,820	4,681
Sussex	573	702	1,709	2,984	1,526	1,155	913	3,594
Union	7,714	5,487	3,668	16,869	6,436	4,141	4,797	15,374
Warren	412	1,282	889	2,583	794	714	2,080	3,588
New Jersey	86,353	93,741	73,635	253,729	86,495	65,959	81,853	234,307

References

Bishop Y., Fienberg M., Stephen E. and Holland P. W. "*Discrete Multivariate Analysis: Theory and Practice*", Cambridge, Massachusetts: The MIT Press, 1975.

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

Cormack R.M. "*The statistics of capture-recapture method*". Oceanographer Marine Biology Annual Review, 6:455-506, 1968.

Doscher Mary-Lynn, and J. Arthur Woodward. "*Estimating the size of sub-populations of heroin users: applications of log-linear models to capture/recapture sampling.*" The International Journal of Addictions, 18(2):167-182, 1983.

French, J.F. "*Prevalence of heroin use in New Jersey*," Unpublished report of Department of Health, Alcohol, Narcotic and Drug Abuse Unit, State of New Jersey, November, 1977(a).

French, J.F. "*Relative heroin prevalence*," Unpublished report of Department of Health, Alcohol, Narcotic and Drug Abuse Unit, State of New Jersey, November 1977(b).

Mammo, Abate, Richard Schadl and Gloria Rodriguez. "*Recent Drug Abuse Variations in Newark*," Community Epidemiology Work Group, in Press, June 1999 Report.

Mammo, Abate. "*Estimating the Number of Drug Abusers in New Jersey Using a Capture-recapture Model*," A Report Submitted to CSAT, November 1, 1995.

Sekar C. G. and Deming W. E. "*On a method of estimating birth and death rates and the extent of registration*", Journal of American Statistical Association. 44: (101-115, 1949.