DOCUMENT RESUME

ED 099 404 TM 004 057

AUTHOR	Beverly, Robert F.
TITLE	A Comparative Analysis of Base Expectancy Tables for Selected Subpopulations of California Youth Authority Wards. Research Report No. 55.
INSTITUTION	California State Dept. of the Youth Authority, Sacramento.
REPORT NO	RR-55
PUB DATE	Dec 68
NOTE	31p.
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	Data Collection; Demography; *Expectancy Tables; *Homogeneous Grouping; *Institutionalized (Persons); Males; Minority Groups; *Predictive Validity; Statistical Analysis
IDENTIFIERS	*California Youth Authority

ABSTRACT

Base expectancy tables developed by the California Youth Authority apply to total population of male wards released on parole by the state. Subpopulations of relatively homogeneous make-up offered the possibility of developing base expectancy tables of greater predictive ability. Multiple regression analysis showed the overall advantages and greater predictive efficiency of using pooled data expectancy tables rather than using subgroup data based on first admissions and readmissions. (SM)

research report no. 55

DEPARTMENT OF THE YOUTH AUTHORITY
STATE OF CALIFORNIA

december, 1968

by
Robert F. Beverly

State of California

RONALD REAGAN
governor

Human Relations Agency

SPENCER WILLIAMS
secretary

Department of the Youth Authority

ALLEN F. BREED
director

GEORGE SALEEBEY
deputy director

KEITH S. GRIFFITHS
Chief of Research

Robert F. Beverly
Senior Social
Research Analyst

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Cont'd)

SUMMARY

Base expectancy tables, developed periodically in the California Youth Authority (CYA) since 1959, have been typically based upon and applied to the *total* population of male wards released to parole supervision within the state. The present study examines several likely *subpopulations* of this total population with the objective of identifying one or more relatively homogeneous groups of wards upon which base expectancy tables might be developed--base expectancy tables which offer promise of greater predictive efficiency for these specifically identified subpopulations than a table based upon the total male population.

Of the subpopulations examined, it initially appeared (on the basis of some fairly impressive data) that those subpopulations defined by the current admission status of the ward (first admission or readmission to the Youth Authority) warranted separate analysis and subsequent construction of differential base expectancy tables. Consequently, three separate base expectancy tables were developed on the basis of construction sample data--one for first admission wards, one for readmission wards, and a third for total admission wards. The predictive efficiency of the first admission table was subsequently compared with that of the total admission table with respect to the first admission wards of an independent sample. Similarly, the predictive efficiency of the readmission table was compared with that of the total admission table as applied to the readmission wards of the same independent sample. As a result of these comparisons, it was concluded that in terms of predictive efficiency, the total admission table was somewhat superior to *both* first admission and readmission tables. It was noted, however, that this particular finding was by no means a necessary one and that there may well exist other subpopulations for which specific base expectancy tables might be more appropriate. However, until such time as these subpopulations are identified, *it would appear that a single base expectancy table will be able to satisfy current Youth Authority needs for this type of instrument.*

For the more technically-oriented reader, there are several methodological changes or innovations which may be of interest. The first of these is the precise specification of the methodological steps followed in

CYA base expectancy table construction as outlined in the Method section on page 4--the essential feature of this innovation being the combination of adjacent base expectancy categories whose associated violation rates are not statistically significantly different from one another ($p > .05$).

The second innovation is the recommended usage of a *preshrunk* base expectancy table in which both the number of base expectancy categories and their associated violation rates are based on validation (or cross-validation) sample data.

The third methodological change involves the usage of what has been termed an "estimated" point biserial correlation coefficient. The characteristics of these estimates in relation to number of base expectancy categories is described and illustrated in Appendix A.

INTRODUCTION

The Division of Research of the California Youth Authority (CYA) has, since early 1959, periodically developed experience (base expectancy) tables relevant to its male population at the time of their release to parole. Based upon base expectancy score,[1] these tables distribute all male wards released to parole within a given period of time among several class intervals or categories, each which specifies the probability of parole violation[2] of construction sample subjects within a specified period of parole exposure--usually 15 months (see Table 3).

There are two related ways in which base expectancy tables have been useful in the evaluation of treatment programs within the CYA. The first of these is in the assessment of the equivalence of comparison groups (e.g., experimental and control) with respect to certain performance-related characteristics as specified in the base expectancy or multiple regression equation. The second is in the statistical "control" of the effects of these characteristics when comparison groups are found to be "non-equivalent" (i.e., to have a differential or disproportionate distribution of wards within the several risk categories of the base expectancy table).

Typically, Youth Authority base expectancy tables have been developed on the basis of data relevant to all male wards released to California parole supervision. However, because of the appreciable diversity of these wards with respect to a number of background characteristics, it has been

[1]Base expectancy score is determined by means of an equation which optimally weights a number of criterion-related variables as determined by multiple regression analysis. These variables include such background characteristics as age at first admission to the CYA; age at release; criminal record, both prior and subsequent to first admission to the CYA; prior school misbehavior, etc.

[2]Parole violation is defined as the removal of a parolee from suspended status by either revocation or discharge to another jurisdiction.

suggested at various times that these tables might be even more predictive of parole performance if they had been differentially developed upon more homogeneous CYA subpopulations, particularly with reference to such variables as age at release, admission status (first admissions as opposed to readmissions), and court of commitment (juvenile vs. criminal). Underlying these suggestions is the implicit assumption of the existence of the effects of interaction--effects which, in the present instance, would be indicative of the differential ability of one or more variables to predict parole outcome for different kinds of parole releasees (e.g., the number of commitments prior to admission to the CYA might well be significantly more predictive of the parole performance of first admission wards than of readmissions).

The purpose of the present study, therefore, is to determine the extent to which those variables predictive of parole outcome in the general male population are differentially predictive for certain specifically defined subpopulations; to develop, if the data are indicative of its feasibility, differential base expectancy equations and tables appropriate to these subpopulations; and to evaluate the predictive efficiency of each subpopulation base expectancy table developed in relationship to that of the total-population base expectancy table.

METHOD

Subjects. The subjects of the present study are the 1963 and 1964 cohorts of male wards released to California parole supervision for whom initial home visit (IHV)data[3] are available. There are 5573 subjects in the 1963 cohort and 6387 in that of 1964. For purposes of statistical

[3]IHV data are obtained by means of a questionnaire completed by the parole agent at the time of the initial home visit. These data are available for approximately 83% of all male wards released to California parole supervision from January 1, 1963 to December 31, 1964.

analysis, it is generally assumed that the calendar-year release cohorts of any defined CYA populations are representative samples of that population. In the present instance, because of its relative recency (certain background characteristics have a tendency to change as a function of time), the 1964 release cohort was selected as the "construction" sample for the development of base expectancy tables, and the 1963 release cohort was selected as the validation sample.

Subpopulations. The subpopulations sampled in this study are defined in terms of the three variables most frequently suggested as providing a sound basis for the establishment of groups for whom specific base expectancy tables might profitably be developed--age at release, admission status, and court of commitment. They are: 1) younger juvenile court first admissions, 2) older juvenile court first admissions, 3) younger juvenile court readmissions, 4) older juvenile court readmissions, 5) older criminal court first admissions, and 6) older criminal court readmissions. Because, with rare exception, the minimum age for criminal court adjudication is 18 years, and because "younger" wards are defined as those who at the time of release to parole are 16 years of age or less (as of last birthday), it is apparent that there can be no "younger" criminal court subpopulations.

Procedures. The initial analysis of the study was directed toward the determination of which of the six subpopulations sampled, or any logical combination thereof,[4] appeared to be sufficiently different from the others to warrant individual multiple regression analysis. The decision was based essentially on the simple inspection of a matrix of chi-squares (Table 1) which, for each of the six subpopulations, indicated the significance of the relationships in the combined 1963 and 1964 samples (cohorts) between each of a number of selected independent variables and the criterion of parole violation or non-violation within 15 months of release.[5] These variables were selected from a larger number of variables on the basis of the statistical significance of their relationship to the criterion ($p < .001$). For a variety of reasons, however, several significantly related variables were excluded from the multiple regression analyses. Attitude toward school was excluded inadvertently (see footnote 6); county of commitment and institution of release were excluded because the inclusion of these variables would have made any analysis of either counties or institutions with respect to the comparability of their wards in terms

[4]Logical combinations of these subpopulations are limited either to the simple addition of all wards who are similar with respect to one dimension (age, admission status, or court of commitment), or to the addition of all wards who are similar with respect to two dimensions (age and admission status, or court and admission status).

[5]Certain between-cohort differences in the criterion-coding of those non-suspended wards discharged from the CYA prior to 15 months of parole exposure resulted in a somewhat higher violation rate for the 1963 cohort than would normally have been the case. The effect, if any, of this higher violation rate on the findings of this report is believed to be negligible.

of probability of parole violation impossible; and commitment offense was excluded as an unreliable variable for several reasons, particularly when juvenile and criminal court cases are combined. The 1963 and 1964 samples were combined in order to provide a sufficiently large number of subjects to minimize the effects of data fractionation occasioned by the division of the total sample into subsamples relevant to the six subpopulations specified above.

Samples (1964 release cohorts) of those subpopulations for whom multiple regression analysis was considered to be appropriate were, along with the total 1964 release cohort, subjected to multiple regression analysis. Each subject was then scored on the basis of both the regression equation applicable to "all" subjects and the regression equation specific to the unique subpopulation to which he belonged. On the basis of these scores, base expectancy tables associated with each multiple regression equation were constructed and the predictive efficiency of each "subpopulation" table was evaluated against that of the total population in an independent sample (1963 release cohort).

Base expectancy table construction. It should be noted that the number of different base expectancy tables that can be constructed from the same distribution of base expectancy scores may be quite large and will be influenced by a variety of factors, the most important of which is probably the use to which the table is to be put. In the present instance, it was intended that each table developed upon construction sample subjects should, as nearly as possible, distribute the subjects of an independent (validation) sample of the same population among a maximum number of differentially predictive base expectancy categories.

The first step toward the achievement of this objective with respect to each given population was to distribute all construction sample subjects by both numerically ordered base expectancy scores, and by violation or non-violation of parole within 15 months of release. Next, by means of trial and error, these distributions of adjacent base expectancy scores were combined into class intervals (base expectancy score categories) in order to arrive at whatever number of "provisional" base expectancy tables satisfied the following conditions:

1. Each class interval of the table contained approximately the same number of subjects;

2. The violation rates associated with successive class intervals of the table showed a consistent increase, or decrease, in magnitude (a reversal in the trend of violation rates did not occur);

Having met these conditions, the table with the maximum number of class intervals possible under which both conditions 1. and 2. obtained was selected. The differences in the violation rates of adjacent class intervals of the selected provisional table were then tested for significance. If the differences between them were not significant ($p > .05$), the class intervals were combined to form a single category, the violation rate of which was subsequently tested against that of its adjacent category. In this fashion, class intervals were collapsed until the violation rate associated with each class interval was significantly different from that of its immediately adjacent neighbor(s), the end product being the construction sample base expectancy table.

RESULTS

Subpopulation Analysis

The results of the initial analysis, as described in the Method section, are shown in Table 1. On the basis of the distribution of chi-squares presented in this table, it may be observed that those subpopulations for which the independent variables listed appear to be decidedly related to parole outcome are: 1) younger juvenile court first admissions, 2) older juvenile court first admissions, and 3) older criminal court first admissions. The single definitive characteristic that the subjects of each one of these subpopulations have in common is that of admission status--they are all first admissions to the CYA. By way of contrast, these same variables are, for the most part, not related to parole outcome for the remaining three subpopulations which, it will be noted, are all readmission groups (previous parole failures). This finding is dramatically illustrated in Table 2 in which the significance of the relationships between the independent variables and violation or non-violation of parole are shown separately for the combined three first admission groups and the combined three readmission groups of Table 1. It will be noted that the independent combining of both first admission and readmission groups as shown in Table 2 allows for the emergence in each group of three independent variables which, by definition, could not be listed in Table 1--age at first admission, age at release, and court of commitment. These

TABLE 1

Significance of Relationships in Terms of Chi-square ($df=1$) Between Seven Dichotomized Independent Variables and Parole Performance for Six Male CYA Subpopulations (Combined 1963 and 1964 Release Cohorts; N=11,939[a])

Independent[b] Variables	Younger Juvenile Court 1st Admissions (N=3570)	Older Juvenile Court 1st Admissions (N=2677)	Younger Juvenile Court Readmissions (N=1053)	Older Juvenile Court Readmissions (N=1798)	Older Criminal Court 1st Admissions (N=2041)	Older Criminal Court Readmissions (N=800)
Number of prior commitments	44.74***	17.64***	0.91	8.23**	23.62***	1.58
Number of prior escapes	16.65***	9.07**	0.78	1.37	6.40*	0.18
School misbehavior	30.51***	11.54***	0.05	1.02	6.84**	0.01
Attitude toward school	16.94***	8.60**	0.66	2.50	6.23*	0.00
Number of offense partners	15.75***	4.59*	0.83	2.86	14.28***	0.35
Number of foster-home placements	20.53***	9.05**	1.45	9.96**	6.81**	1.00
Psychological evaluation	5.74*	5.28*	0.31	1.06	13.41***	0.90

[a]The total number of subjects in the analyses performed in Tables 1 and 2 is slightly less than the total number of subjects in the combined cohorts as subsequently analyzed (N=11,960).

[b]Variables are dichotomized as indicated in Tables 3A-5A.

* $p < .05$; ** $p < .01$; and *** $p < .001$.

TABLE 2

Significance of Relationships in Terms of Chi-square ($df=1$) Between Ten Independent Variables and Parole Performance for First Admissions and Readmissions to the CYA (Combined 1963 and 1964 Release Cohorts; N=11,939)

Independent Variables	First Admissions (N=8288)	Readmissions (N=3651)
Number of prior commitments	57.45*	0.71
Number of prior escapes	39.72*	0.29
School misbehavior	95.07*	1.52
Attitude toward school	29.29*	1.03
Number of offense partners	42.41*	5.48
Number of foster-home placements	63.68*	21.66*
Psychological evaluation	40.51*	2.16
Court of commitment	258.86*	98.22*
Age at first admission	423.53*	111.33*
Age at release to parole	444.18*	131.15*

*$p < .001$

three variables, as is the case with number of foster home placements, are highly significant ($p < .001$) for both first admission and readmission groups.

Multiple Regression Analyses

The finding illustrated in Tables 1 and 2 that a number of independent variables are differentially predictive of parole outcome for first admissions and readmissions to the CYA resulted in three separate multiple regression analyses of the 1964 release cohort--one each for first admissions, readmissions, and total admissions (first admissions and readmissions combined). With the exception of attitude toward school,[6] only those variables of Table 2 which are listed as being significant ($p < .001$) in the appropriate admission status column were used in the first admission and readmission regression analyses--a total of nine variables for the former and four for the latter. Although not shown, the same nine variables of Table 2 (excluding attitude toward school) found to be significant for the first admission sample were, along with admission status classification, found to be significant for the total admission sample.

Tables 3-5 present the base expectancy formulae (multiple regression equations) and base expectancy tables for first admissions, readmissions and total admissions respectively. It will be noted that each formula retains only those of the original variables which multiple regression analysis found to contribute uniquely to parole performance variance in the designated population ($p < .05$). In the case of first admissions, five of nine original variables are retained in the formula. Three of four and eight of ten original variables are retained in the readmission and total admission equations respectively. It will also be observed that the number of categories of each of the three base expectancy tables is directly related to the number of independent variables in the formula upon which it is based.

BE Table Validity (First Admissions)

Table 6 compares the predictive ability of the first admission base expectancy table (Table 3) with that of the total admission base expectancy

[6] The author inexplicably omitted this variable from both the first admission and total admission multiple regression analyses. There should be no appreciable loss, however, as the highly significant relationship of "attitude toward school" to the more powerfully predictive item of "school misbehavior" (unpublished data by Martin J. Molof of the Division of Research) makes it unlikely that its *unique* contribution to either of the prediction equations, if any, would be more than minimal.

TABLE 3

Base Expectancy Formula and Base Expectancy Table for Male CYA First Admission Wards Based on the 1964 Parole Release Cohort

A. BASE EXPECTANCY FORMULA

1. Constant 192.2

2. Age at first admission:

   19^{+} = 383.9   17 = 332.4   15 = 123.4 13^{-} = 0
   18 = 352.5   16 = 237.8   14 = 95.6 .... _____

3. Number of prior commitments:

   none = 96.2,   1 or more = 0 .... _____

4. School misbehavior:

   none = 66.6,   some = 0 .... _____

5. Number of prior escapes:

   none = 66.5   1 or more = 0 .... _____

6. Number of offense partners:

   1 or more = 32.4,   none = 0 .... _____

Total (BE Score)# =

B. BASE EXPECTANCY TABLE (15 months of parole follow-up)

Category	BE Score	Total	V*	N-V*	p_V*
1	738^{+}	632	137	495	.217
2	611-737	1207	420	787	.348
3	529-610	540	221	319	.409
4	482-528	623	294	329	.472
5	387-481	613	342	271	.558
6	192-386	585	399	186	.682
		4200	1813	2387	(.432)

est. r_{pb} = .28Ψ

#Round to nearest whole number.

*V = no. of violators, N-V = no. of non-violators, and p_V = proportion of V.

ΨEstimated point biserial correlation coefficient (see Appendix A).

TABLE 4

Base Expectancy Formula and Base Expectancy Table for Male CYA Readmission Wards Based on the 1964 Parole Release Cohort

A. BASE EXPECTANCY FORMULA

1. Constant 239.9

2. Age at release:

19^+ = 229.6, 18 = 146.9, 17 = 144.2, 16^- = 0 ____

3. Court of commitment:

criminal = 92.0, juvenile = 0 .... ____

4. Foster home placement:

none = 51.8, 1^+ = 0 .... ____

Total (BE Score)[#]

B. BASE EXPECTANCY TABLE (15 months of parole follow-up)

Category	BE Score	Total	V*	N-V*	P_V*
1	527^+	484	191	293	.395
2	383-526	1092	606	486	.555
3	240-382	611	442	169	.723
		2187	1239	948	(.567)

est. r_{pb} = .23[Ψ]

[#] Round to nearest whole number.

[*] V = no. of violators, N-V = no. of non-violators, and P_V = proportion of V.

[Ψ] Estimated point biserial correlation coefficient (see Appendix A).

TABLE 5

Base Expectancy Formula and Base Expectancy Table for All Male CYA Wards (Total Admissions) Based on the 1964 Parole Release Cohort

A. BASE EXPECTANCY FORMULA

1. Constant 119.3
2. Age at release:
   19^+ = 232.6   17 = 176.4   15 = 74.5
   18 = 183.9   16 = 90.1   14 = 0 .... _____
3. Age at first admission:
   18^+ = 128.1,   17 = 111.0,   16 = 62.7,   15 = 0 .... _____
4. CYA Admission status:
   first admission = 125.3,   readmission = 0 .... _____
5. Number of prior commitments:
   none = 79.2, 1^+ = 0 .... _____
6. Number of foster home placements:
   none = 39.9, 1^+ + 0 .... _____
7. School misbehavior:
   none = 37.6   some = 0 .... _____
8. Psychological evaluation:
   no = 27.7,   yes = 0 .... _____
9. Number of offense partners:
   1^+ = 27.3,   none = 0 .... _____

Total (BE Score)[#] =

B. BASE EXPECTANCY TABLE (15 months of parole follow-up)

Category	BE Score	Total	V*	N-V*	P_V*
1	751^+	451	92	359	.204
2	665-750	898	283	615	.315
3	548-664	1385	557	828	.402
4	481-547	916	417	499	.455
5	419-480	887	475	412	.536
6	363-418	967	589	378	.609
7	312-362	441	305	136	.692
8	119-311	442	334	108	.756
		6387	3052	3335	(.478)

est. r_{pb} = .30[Ψ]

[#] Round to nearest whole number.

[*] V = no. of violators, N-V = no. of non-violators and P_V = proportion of V.

[Ψ] Estimated point biserial correlation coefficient (see Appendix A).

table (Table 5) when each is applied to the first admission subjects of a validation sample (1963 release cohort). This comparison is made in terms of point biserial correlation coefficient *estimates* (see Appendix A). Since Table 1 of Appendix A shows no reduction in the magnitude of point biserial correlation coefficient estimates when the number of base expectancy categories is reduced from eight to six, it may be safely assumed that the r_{pb}s underlying the first admission and total admission tables, as expressed by the estimated r_{pb}s of .28 and .29 respectively, are practically identical. Therefore, the advantage, if any, would appear to reside in the total admission table because of its ability to differentiate the validation sample into eight distinct categories rather than the six categories of the first admission table.

TABLE 6

Comparison of Construction Sample First and Total Admission Base Expectancy Tables as Applied to a Validation Sample of First Admission Wards (1963 release cohort)

First Admission Table

BE Score	Total	V*	N-V*	p_V*
738+	722	182	540	.252
611-737	1172	433	739	.369
529-610	452	188	264	.416
482-528	542	286	256	.528
387-481	639	367	272	.574
0-386	575	387	188	.673
	4102	1843	2259	(.449)

est. r_{pb} = .28[Ψ]

Total Admission Table

BE Score	Total	V*	N-V*	p_V*
751+	520	134	386	.258
665-750	905	282	623	.312
548-664	917	386	531	.421
481-547	459	232	227	.505
419-480	485	264	221	.544
363-418	512	327	185	.639
312-362	226	154	72	.681
119-311	78	64	14	.821
	4102	1843	2259	(.449)

est. r_{pb} = .29[Ψ]

*V = no. of violators, N-V = no. of non-violators and p_V = proportion of V.

ΨEstimated point biserial correlation coefficient (see Appendix A).

It should be noted that the estimated point biserial correlation coefficients associated with the construction sample tables (Tables 3-5) are in reality *multiple* correlation coefficients and as such, despite the large numbers of subjects in relation to the numbers of variables, might show a tendency toward shrinkage in magnitude when the multiple regression equations upon which prediction scores are based are applied to independent (validation) samples.

In the present instance, the shrinkage in the estimated r_{pb} from Table 3 to Table 6 (the first admission table only) is *nil*, the coefficient being .28 in each case.[7] It is suggested, however, that yet another criterion of shrinkage might be whether or not the proportions of violators associated with adjacent categories of Table 6 are significantly different, one from the other ($p < .05$). If they are not, it is proposed that they be combined until the differences between the proportions of violators in all adjacent categories attain statistical significance ($p < .05$), as described in the Method section and carried out in the construction of Table 3.

The results of testing for differences, and the combining of subjects of non-significantly different adjacent categories of Table 6, resulted in the first admission and total admission tables shown in Table 7. The very slight drop in the estimated point biserial coefficients of each table with the diminution in the number of base expectancy categories will be observed--.27 from .28 for first admissions and .28 from .29 for total admissions. It will also be noted that the six-category first admission table of Table 6

[7]This same type of comparison (i.e., the comparison of the results of the application of the total admission table of the construction sample to only first admission wards of both construction and validation samples) cannot be made because construction sample data relevant to total admissions do not identify the subjects as to admission status.

TABLE 7

Comparison of First and Total Admission Base Expectancy Tables of Table 6 Subsequent to Combination of Non-significantly Different Categories (p > .05)

First Admission Table

BE Score	Total	V*	N-V*	p_V*
738+	722	182	540	.252
529-737	1624	621	1003	.382
387-528	1181	653	528	.553
192-386	575	387	188	.673
	4102	1843	2259	(.449)

est. r_{pb} = .27[Ψ]

Total Admission Table

BE Score	Total	V*	N-V*	p_V*
751+	520	134	386	.258
665-750	905	282	623	.312
548-664	917	386	531	.421
419-547	944	496	448	.525
312-418	738	481	257	.652
119-311	78	64	14	.821
	4102	1843	2259	(.449)

est. r_{pb} = .28[Ψ]

*V = no. of violators, N-V = no. of non-violators, and p_V = proportion of V.

[Ψ]Estimated point biserial correlation coefficient (see Appendix A).

was reduced to four categories with the eight-category total admission table being reduced to six categories. This demonstrable shrinkage in the number of base expectancy categories (as operationally defined) in the validation sample would seem to indicate the methodological superiority of developing, if possible, a *preshrunk* base expectancy table.

BE Table Validity (Readmissions)

Table 8 compares the predictive ability of the readmission base expectancy table (Table 4) with that of the total admission base expectancy table (Table 5) when each is applied to the readmission subjects of a validation sample (1963 release cohort). As was the case with first admissions, this comparison shows the estimates of the point biserial correlation coefficients to be practically identical. This essential equi-

TABLE 8

Comparison of Construction Sample Readmission and Total Admission Base Expectancy Tables as Applied to a Validation Sample of Readmission Wards (1963 release cohort)

Readmission Table

BE Score	Total	V*	N-V*	p_V*
527+	307	141	166	.459
383-526	719	408	311	.567
240-382	445	322	123	.724
	1471	871	600	(.592)

est. r_{pb} = .19ᵠ

Total Admission Table

BE Score	Total	V*	N-V*	p_V*
751+	--	--	--	
665-750	27	13	14	.481
548-664	255	110	145	.431
481-547	220	116	104	.527
419-480	266	148	118	.556
363-418	283	183	100	.647
312-362	193	137	56	.710
119-311	227	164	63	.722
	1471	871	600	(.592)

est. r_{pb} = .20ᵠ

*V = no. of violators, N-V = no. of non-violators, and p_V = proportion of V.

ᵠEstimated point biserial correlation coefficient (see Appendix A).

valence of coefficients is not seriously threatened, even by the tendency, as shown by the data presented in Table 1 of Appendix A, for the absolute value of the point biserial estimates to decrease by a maximum of .02 when the number of categories of the base expectancy table is decreased from seven to three. Any compensating adjustment of either table to take this into account still leaves the two estimates separated by an absolute value of only .01. The advantage once again would therefore appear to reside with the total admission table because of its ability to differentiate the validation sample into six distinct categories (omitting the 665-750 class interval of only 27 cases), as opposed to the three categories of the readmission table.

In the present case, the shrinkage in the point biserial coefficient estimate from the construction sample (Table 4) to the validation sample (Table 8) was from .23 to .19.[8] It will be remembered that there was no shrinkage in the case of the first admission table. When shrinkage is assessed by the combining of those adjacent categories of Table 8 whose proportions of violators are not significantly different from one another, Table 9 is the result. Here, in comparing the readmission table with the total admission table, the estimated point biserials are once again essentially equivalent,[9] and the total admission table provides four distinct

TABLE 9

Comparison of Readmission and Total Admission Base Expectancy Tables of Table 8 Subsequent to Combination of Non-significantly Different Categories ($p > .05$)

Readmission Table

BE Score	Total	V*	N-V*	p_V*
527+	307	141	166	.459
383-526	719	408	311	.567
240-382	445	322	123	.724
	1471	871	600	(.592)

est. $r_{pb} = .19^{\Psi}$

Total Admission Table

BE Score	Total	V*	N-V*	p_V*
548+	282	123	159	.436
419-547	486	264	222	.543
363-418	283	183	100	.647
119-362	420	301	119	.717
	1471	871	600	(.592)

est. $r_{pb} = .21^{\Psi}$

*V = no. of violators, N-V = no. of non-violators, and p_V = proportion of V.

$^{\Psi}$Estimated point biserial correlation coefficient (see Appendix A).

[8] See footnote 7, substituting "to only readmission wards" for "to only first admission wards" in the parenthetical phrase.

[9] Interestingly enough, the point biserial estimate of the four-category total admission table of Table 9 *increased* to .21 from an estimate of .20 in the seven-category total admission table of Table 8. The slightly lower figure of Table 8 was probably the result of the reversal of the violation rate of the 27 subjects in the highest-score category (class interval).

categories as opposed to only three for the readmission table (in terms of numbers of categories, the three-category readmission table had very little room for shrinkage).

BE Table Validity (Total Admissions)

As previously stated, it is not possible to ascertain the shrinkage of the total admission construction sample table (Table 5) when applied solely to either first admissions or readmissions in both construction and validation samples (see footnote 7). However, it is possible to determine the categorical shrinkage of this table when applied to total admissions in a manner identical to that previously employed with the component admission groups of validation sample first admissions and readmissions. Table 10 shows the result of the application of Table 5 to a validation sample

TABLE 10

Total Admission Construction Sample Table (Table 5) as Applied to a Validation Sample of Total Admissions Wards (Table A) and Table A Subsequent to Combination of Non-significantly Different Categories (Table B)

Table A

BE Score	Total	V*	N-V*	P_V*
751+	520	134	386	.258
665-750	932	295	637	.317
548-664	1172	496	676	.423
481-547	679	348	331	.513
419-480	751	412	339	.549
363-418	795	510	285	.642
312-362	419	291	128	.695
119-311	305	228	77	.748
	5573	2714	2859	(.487)

est. r_{pb} = .29$^{\Psi}$

Table B

BE Score	Total	V*	N-V*	P_V*
751+	520	134	386	.258
665-750	932	295	637	.317
548-664	1172	496	676	.423
419-547	1430	760	670	.531
363-418	795	510	285	.642
119-362	724	519	205	.717
	5573	2714	2859	(.487)

est. r_{pb} = .29$^{\Psi}$

*V = no. of violators, N-V = no. of non-violators and P_V = proportion of V.

$^{\Psi}$Estimated point biserial correlation coefficient (see Appendix A).

(1963 release cohort) of total admissions and the consequence of the combination of adjacent categories with non-significantly different proportions of violators (p > .05).

A particularly interesting phenomenon related to the total admission base expectancy formula is readily apparent when Table 6 is placed in juxtaposition with Table 8. These tables show the result of the differential application of the total admission formula to first admissions and readmissions in the validation sample (1963 release cohort). For the convenience of the reader, these eight-category tables are shown together in Table 11. With the exceptions of the first two class intervals (populated by only 27 readmissions) and the last class interval, it will be noted that the proportions of violators of the class intervals of both tables are remarkably

TABLE 11

Total Admission Base Expectancy Table (Formula) as Applied to First Admissions (Table 6) and Readmissions (Table 8) in a Validation Sample (1963 release cohort)

First Admission Table

BE Score	Total	V*	N-V*	p_V*
751+	520	134	386	.258
665-750	905	282	623	.312
548-664	917	386	531	.421
481-547	459	232	227	.505
419-480	485	264	221	.544
363-418	512	327	185	.639
312-362	226	154	72	.681
119-311	78	64	14	.821
	4102	1843	2259	(.449)

est. r_{pb} = .29ᵠ

Readmission Table

BE Score	Total	V*	N-V*	p_V*
751+	--	--	--	--
665-750	27	13	14	.481
548-664	255	110	145	.431
481-547	220	116	104	.527
419-480	266	148	118	.556
363-418	283	183	100	.647
312-362	193	137	56	.710
119-311	227	164	63	.722
	1471	871	600	(.592)

est. r_{pb} = .20ᵠ

*V = no. of violators, N-V = no. of non-violators and p_V = proportion of V.

ᵠEstimated point biserial correlation coefficient (see Appendix A).

similar--the respective violation rates of the middle five categories of first admissions and readmissions being .421 and .431, 505 and .527, .544 and .556, .639 and .647, and .681 and .710. For the most part, it would therefore appear that base expectancy scores obtained by first admission subjects are associated with the same probability of parole violation as those of readmissions with similar base expectancy scores; or, stated differently, a base expectancy score obtained by the addition of scores (weights) associated with one particular combination of background characteristics is essentially equivalent (in terms of associated violation rates) to the same base expectancy score obtained by an entirely different combination of background characteristics.

DISCUSSION AND CONCLUSIONS

As indicated by the magnitude of the chi-squares shown early in the Results section (Table 2), there appeared to be little question as to the differential effectiveness of a large proportion of the independent variables listed with respect to their ability to predict parole performance (violation or non-violation of parole within fifteen months of release to such status) for first admission and readmission male wards. On this basis, separate construction sample multiple regression analyses of these subpopulations (as well as that of the total population) were performed and three base expectancy tables constructed. Contrary to expectation, however, it was found that when these first admission and readmission tables were applied to validation samples of first admission and readmission subjects, their respective predictive efficiencies (in terms of magnitude of estimated point biserial correlation coefficients) were essentially the same (slightly lower) as those obtained by use of the base expectancy table derived from analysis of the total population (first admissions and readmissions combined). Furthermore, in terms of specified procedural criteria, it was found in each comparative instance that the total admission base expectancy table, as applied separately to both first admissions and readmissions in the validation sample, differentiated the subjects into a larger number of base expectancy score categories (class intervals), each with a significantly different violation rate than that of its immediately adjacent neighbor(s).

A greater number of base expectancy categories with distinct violation rates allows for the better statistical control of differences in the background characteristics of experimental and comparison group subjects which frequently occur despite rigorous adherence to standard procedures of either random assignment to, or the matching of subjects for, differential treatment groups.

Hypothetically, a further advantage of a greater number of base expectancy categories occurs in those situations in which administrative decisions call for the selection of particularly good or poor risk wards for a given institutional or community program. This potential advantage is more pronounced in the case of first admission wards than readmission wards. With respect to the latter, the difference in the ability of the readmission and total admission tables to discriminate among readmissions is essentially negligible (Table 9).

In the case of first admissions, however, the total admission table of Table 7 demonstrates the two advantages over the first admission table specified above. At the favorable end of the continuum, it enables one to select out 1425 wards (the top two categories) with a violation rate of only 29%, while the first admission table permits either the selection of 722 wards with a violation of 25% or 2346 wards (the top two categories combined) with a violation rate of 34%. At the opposite end of the continuum, the advantage of the total admission table resides in its ability to select wards of extremely poor risk quality, ones whose rate of violation is 82%. This ability to categorize extremely poor risk subjects is one not possessed by the first admission table. Perhaps a more direct way of remarking upon the advantage of the six-category total admission table as opposed to the four-category first admission table of Table 7 is to point out that although the categories of the latter can be closely approximated in terms of violation rate by the combination of the appropriate pairs of total admission table categories, the reverse is not true--there exist two unique categories of the total admission table whose associated violation rates can in no way be approached by the first admission table categories. Specifically, these total admission table categories are those whose violation rates are .312 and .821.

In view of the seemingly contradictory data presented in Table 2, one might well ponder the demonstrated ability of a regression equation based upon a heterogeneous group of total admissions to better differentiate among both first admission and readmission subjects than equations based exclusively upon these homogeneous groups themselves. This paradox may best be understood in terms of conclusions drawn from the following:

1. The magnitude of the differences in chi-squares shown for first admissions and readmissions in Table 2 may be *partially* attributed to the differences in the numbers of subjects in those respective groups. In the combined release cohorts of 1963 and 1964 these are 8288 first admissions

as opposed to only 3651 readmissions, or 2.27 times as many of the former as of the latter. This explanation results from the fact that with the same absolute differences in violation rate between two groups, chi-square will vary directly as a function of the number of subjects in each group.

2. When tested for significance by means of a special type of analysis of variance described by Rao (1962),[10] it is found that despite the magnitude of the differences in first admission and readmission chi-squares shown for the ten variables listed in Table 2, there are only three variables in which the relationships between the violation rates of the first admission and readmission dichotomies are significantly different ($p < .01$) from each other. The three variables in which these significant interaction effects are found are: number of commitments prior to commitment to the CYA, number of escapes prior to CYA commitment, and school misbehavior.

3. With respect to the remaining seven variables listed in Table 2 (those not specified in 1. above), each of which exhibited no significant interaction effect as a function of admission status, the direction of the relationship between the violation rates of each of the dichotomies is the same, but considerably less pronounced in the case of readmissions than first admissions. In most cases, a larger chi-square for total admissions than for first admissions alone is the result.[10]

Consideration of the findings presented in this study has led to the specific conclusion that despite the fact that some variables are significantly more predictive of successful parole performance for first admission wards than for readmissions and vice versa, there are overall advantages to be gained in terms of comparative predictive efficiency for a base

[10]Data not shown.

expectancy table derived from the multiple regression analysis of pooled admission status subgroup data as opposed to base expectancy tables derived from the multiple regression analysis of homogeneous subgroup data peculiar to first admissions and readmissions individually.

Unfortunately, this conclusion cannot be stated in a generalizable form. This is because the findings of this type of study are undoubtedly related to the relative importance to prediction of those variables (in terms of beta weights) in which significant interaction effects are found and those in which they are not found. In the present instance, the advantages to prediction of pooled first admission and readmission data in those variables with insignificant interaction effects outweighs the disadvantages to prediction in those variables with significant interaction effects. In another case, the reverse might well obtain. However, until such time as independent variables as predictive of parole outcome as age are found, this result is not very likely.

REFERENCES

Beverly, R. F. Base expectancies and the Initial Home Visit Research Schedule. *Calif. Dept. Youth Authority, Div. of Research Report, 1964, No. 37.*

Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, (Fourth Edition), 1965.

Rao, C. R. *Advanced statistical methods in biometric research.* New York: John Wiley & Sons, Inc., 1962.

APPENDIX A

Estimates of Point Biserial Correlation Coefficients

Guilford (1965) has shown mathematically that the point biserial correlation coefficient formula is but a special case of the Pearson product moment formula in which the independent variable X is both continuous and continuously measured and Y is a genuine dichotomy with point values of 0 and 1. In the present instance, with base expectancy score designated as the X variable and parole performance (violation or non-violation of parole within a specified time period) designated as Y, it is found that the raw score product moment formula yields a correlation coefficient identical with that obtained by use of the point biserial formula.

Furthermore, if subjects are grouped by means of base expectancy score into a variable number (n) of successive class intervals (categories) containing approximately equal numbers of subjects, and if the score values of the categories are arbitrarily set so as to progress from 1 to n, it is found that with base expectancy tables of four or more categories ($n \geq 4$), correlation coefficients obtained by use of the raw score product moment formula only *very slightly* underestimate that obtained when X is treated as a fully continuous and continuously measured variable. With n smaller than 4, however, the magnitude of the underestimation tends to increase more steeply. This decreasing relationship may be observed in the coefficients associated with the different size base expectancy tables presented below in Table 1.

The point biserial coefficient estimates of Table 1 were calculated by application of the raw score product moment formula to two sets of base

APPENDIX A (Cont'd)

expectancy tables of variable size as presented in a previous publication by Beverly (1964) and to one set of tables constructed essentially from the eight-category table presented in Table 5B. Each set is associated with a different multiple regression equation and contains base expectancy tables of 24, 12, 8, 6, 4, 3, and 2 categories in the case of Equations #1 and #2 and 8, 4, 3, and 2 categories in the case of Equation #3. It will be noted that the estimates of r_{pb} for Equation #1 and #3 which are obtained from base expectancy tables with four or more categories are only slightly under their conventionally calculated values--the three category estimates being the first to decrease to any considerable extent (more than an absolute value of .01). For Equation #2, the underestimates are similarly slight with all tables of *three* or more categories.

Table 1

Estimates of Point Biserial Coefficients of Correlation between Base Expectancy Score and Violation of Parole as Related to Number of Base Expectancy Table Categories of Three Multiple Regression Equations

Number of Expectancy Table Categories	Equation #1* (r_{pb} = .31)	Equation #2* r_{pb} = .27)	Equation #3* r_{pb} = .30)
	r_{pb} ESTIMATES		
24	.30	.27	--
12	.30	.27	--
8	.30	.27	.30
6	.30	.27	--
4	.30	.26	.29
3	.28	.26	.27
2	.27	.23	.25

*Equations #1 and #2 from Beverly (1964); Equation #3 from Table 5A, page 10 of the present report.