THE STATISTICAL ASSOCIATION BETWEEN PAST AND FUTURE ACCIDENTS AND VIOLATIONS

By
J. Richard Stewart
B. J. Campbell

The University of North Carolina
Highway Safety Research Center
Chapel Hill, North Carolina

December 1972

ABSTRACT

The predictability of future accidents in terms of past violations or past accidents is investigated by observing a four-year history of accident and violation records of North Carolina drivers. The four-year driving history was divided into two adjacent time intervals, and the relationships were determined between accidents and violations. Other factors considered are driver age and the relative lengths of the two time periods.

The results show that past accidents are somewhat better predictors of future accidents than are past violations, in the sense of identifying high risk driver groups. In either case, however, the vast majority of all accidents occurring in a period of time (one, two or three years) involve drivers having no accidents or violations in the previous period.

Driver age seems to have little effect with respect to the relationships between past and future driving performance with the exception that the performance of the youngest group of drivers seems to be slightly more unpredictable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

The authors wish to express their appreciation to the Department of Motor Vehicles and to the Governor's Highway Safety Program, the cooperation of which made it possible to compile and analyze these data. They also wish to thank Gary G. Koch for his assistance in using the weighted least square procedures described in the appendix.

INTRODUCTION

It has long been recognized that the future driving performance of an individual driver is somewhat related to his past driving record (e.g., drivers with more than average numbers of accidents and violations in the past tend to have more than average numbers of accidents and violations in the future).

The relationship is weak, however, and past driving records account for a very small proportion of the subsequent accident and violation variance.

This study is addressed to the investigation of the relationships between accidents and violations over two consecutive time intervals.

The basic data for this study consist of accident and violation counts for each of four years beginning with December 1966 and continuing through November 1970, for 2,502,240 North Carolina drivers who were at least 22 years old at the end of the study period. Additional data were collected over the last two years of the study period for 368,025 drivers who were 21 years old at the end of the study period. These data were compiled by a search of the entire North Carolina Department of Motor Vehicles (DMV) Driver License File, which contains the official driving records of all North Carolina drivers.

The four-year study period was divided into two two-year intervals, and the relationships between occurrences of accidents and violations in the two time intervals were investigated for various driver age groups. For the youngest drivers only, two years of data were compiled—divided into two one-year periods.

ANALYSIS AND RESULTS

For analyses the data were arrayed into two-way tables where rows are the occurrences (accidents or violations) in the *first* time period, and columns are the occurrences in the second time period. Row and column sums, measures of association or correlation, and other summary information were then computed for each table.

Table 1 shows accidents in the second time period (two years) versus violations in the first two-year period for drivers old enough to have a four-year driving record. Table 2 shows accidents in the second period versus accidents in the first two-year period. Thus Table 1 shows the predictive power of past violations relative to future accidents and Table 2 shows the same for past accidents relative to future accidents.

In both tables, numbers of violations and accidents are recorded as 0, 1, . . . , 6 and 7 or more. Measures of association or correlation (namely Goodman's G, Spearman's ρ , and Kendall's τ) have been computed and are shown with the tables. Tables 1 and 2 show that the correlations are quite low. However past violations have a slightly higher correlation with later accidents than past accidents have with future accidents. This finding agrees with Peck, McBride, and Coppin (1971).

Despite this overall difference in correlation slightly favoring violations, there is another way of looking at the data that favors accidents as a predictor. Tables 1-A and 2-A show certain summary measures computed from the data in Tables 1 and 2, respectively. Thus, for example, the fifth row shows the proportion of drivers having at least one accident during the second two-year period as a function of the number of violations (Table 1-A) or accidents (Table 2-A) they experienced in the first two-year period. Table 1-A shows that the proportion having future accidents ranges from approximately 10 percent for drivers having no violations during the first period, up to nearly 30 percent for drivers having four violations in the first period. The values level off at approximately the 30 percent level for drivers with more than four violations in the first period.

Thus, of the 674 drivers in the state who had seven or more violations in the first time period, 30 percent had at least one accident in the second time period, but 70 percent had none.

Table 2-A shows this proportion again beginning at about .10 for drivers having no *accidents* in the first period. In this case, however, the values rise steadily to .70 for drivers, with seven or more accidents in the first period. That is, of the ten drivers with seven or more past accidents (two years), 70 percent had at least one accident in the following two years. Plots of these proportions are shown in Figures 1 and 2 as functions of prior violations and prior accidents respectively.

Since there were more drivers experiencing a given number of violations in the first period than experiencing the same number of accidents in the first period it might seem that if more columns were included in Table 1-A, the proportion of drivers with one or more accidents in the second period might increase beyond .30 for some greater number of violations. However, this was still not the case even when categories of 7, 8 and 9+ violations were included. Another possible explanation for this difference between Figure 1 and Figure 2 in the proportion having one or more accidents in the second period might be that drivers having a large number of violations in a short period of time may tend to be removed from the driving population by state action and, hence, not be available to experience accidents in a later period, while, on the other hand, drivers may not be routinely removed from the driving population as a function of accidents alone.

Table 1. Accident distributions over a two-year period as a function of violations in a prior two-year period.

		Accidents—second two-year period								Row Total
		0	1	2	3	4	5	6	7+	
Violations—first two-year period	0	1,890,951	185,069	18,451	2,094	301	45	19	5	2,096,935
	1	247,266	43,379	6,759	1,012	181	39	5	4	298,645
	2	56,457	13,407	2,693	532	107	17	3	0	73,216
	3	15,862	4,647	1,078	247	57	13	3	0	21,907
	4	5,073	1,599	416	105	19	7	2	3	7,224
	5	1,848	570	149	24	5	1	0	0	2,579
	6	729	216	77	10	6	2	0	2	1042
	7+	474	154	38	4	4	0	0	0	674
COLUMN TOTAL		2,218,660	249,041	29,661	4,028	680	124	32	14	2,502,240

G = .3604 Sd = .0019

ρ = .1130

τ = .0265 Sd = .0002

Table 1-A. Summary measures derived from Table 1.

No. of Violations During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of violations during first time period	2,096,935	298,645	73,216	21,907	7,224	2,579	1,042	674	2,502,240
Proportion of total no. of drivers	.8380	.1194	.0293	.0088	.0029	.0010	.0004	.0003	1.00
No. of accidents in second period by drivers having this no. of violations	229,831	60,910	20,920	7,855	2,890	965	448	258	324,077
Proportion of total accidents by drivers having this no. of violations	.7092	.1879	.0646	.0242	.0089	.0030	.0014	.0008	*
Proportion of drivers with this no. of violations having at least one accident during second period	.0982	.1720	.2289	.2759	.2978	.2884	.3004	.2967	.1133
No. of drivers having at least this no. of violations	2,502,240	405,305	106,660	33,444	11,537	4,313	1,716	674	*
No. of accidents in second period by drivers having at least this no. of violations	324,077	94,246	33,336	12,416	4,561	1,671	706	258	*
Proportion of all drivers having at least this no. of violations	1.00	.1620	.0426	.0134	.0046	.0017	.0007	.0003	*
Proportion of all accidents by drivers having at least this no. of violations	1.00	.2908	.1029	.0383	.0144	.0052	.0022	.0008	*
Average no. of accidents in second period by drivers having this no. of violations	.1096	.2040	.2857	.3586	.4001	.3716	.4299	.3828	.1295

*No additional information provided by this cell

Table 2. Accident distribution over a two-year period as a function of accidents in a prior two-year period.

Accidents—first two-year period	Accidents—second two-year period								Row Total
	0	1	2	3	4	5	6	7+	
0	2,002,577	206,778	22,080	2,639	406	71	19	7	2,234,577
1	192,684	35,363	5,842	986	172	29	3	1	235,080
2	20,467	5,694	1,369	300	65	18	5	1	27,919
3	2,546	1,008	284	81	24	5	1	4	3,953
4	321	162	73	14	11	1	2	0	584
5	55	26	7	7	1	0	2	1	99
6	7	6	4	1	0	0	0	0	18
7+	3	4	2	0	1	0	0	0	10
COLUMN TOTAL	2,218,660	249,041	29,661	4,028	680	124	32	14	2,502,240

$G = .3441$ $Sd = .0023$

$\rho = .0893$

$\tau = .0175$ $Sd = .0002$

Table 2-A. Summary measures derived from Table 2.

No. of Accidents During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of accidents during first period	2,234,577	235,080	27,919	3,953	584	99	18	10	2,502,240
Proportion of total no. of drivers	.8930	.0939	.0112	.0016	.0002	.0000	.0001	.0000	*
No. of accidents in second period by drivers having this no. of accidents	260,997	50,863	9,719	1,974	411	84	17	12	324,077
Proportion of total accidents by drivers having this no. accidents	.8054	.1569	.0300	.0061	.0013	.0003	.0001	.0000	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.1038	.1803	.2669	.3559	.4503	.4444	.6111	.7000	.1133
No. of drivers having at least this no. of accidents in first period	2,502,240	267,663	32,583	4,664	711	127	28	10	*
No. of accidents in second period by drivers having at least this no. of accidents	324,077	63,080	12,217	2,498	524	113	29	12	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.1070	.0130	.0019	.0003	.0001	.0000	.0000	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.1946	.0377	.0077	.0016	.0003	.0001	.0000	*
Average no. of accidents in second period by drivers having this no. of accidents	.1168	.2164	.3481	.4994	.7038	.8485	.9444	1.20	.1295

*No additional information provided by this cell

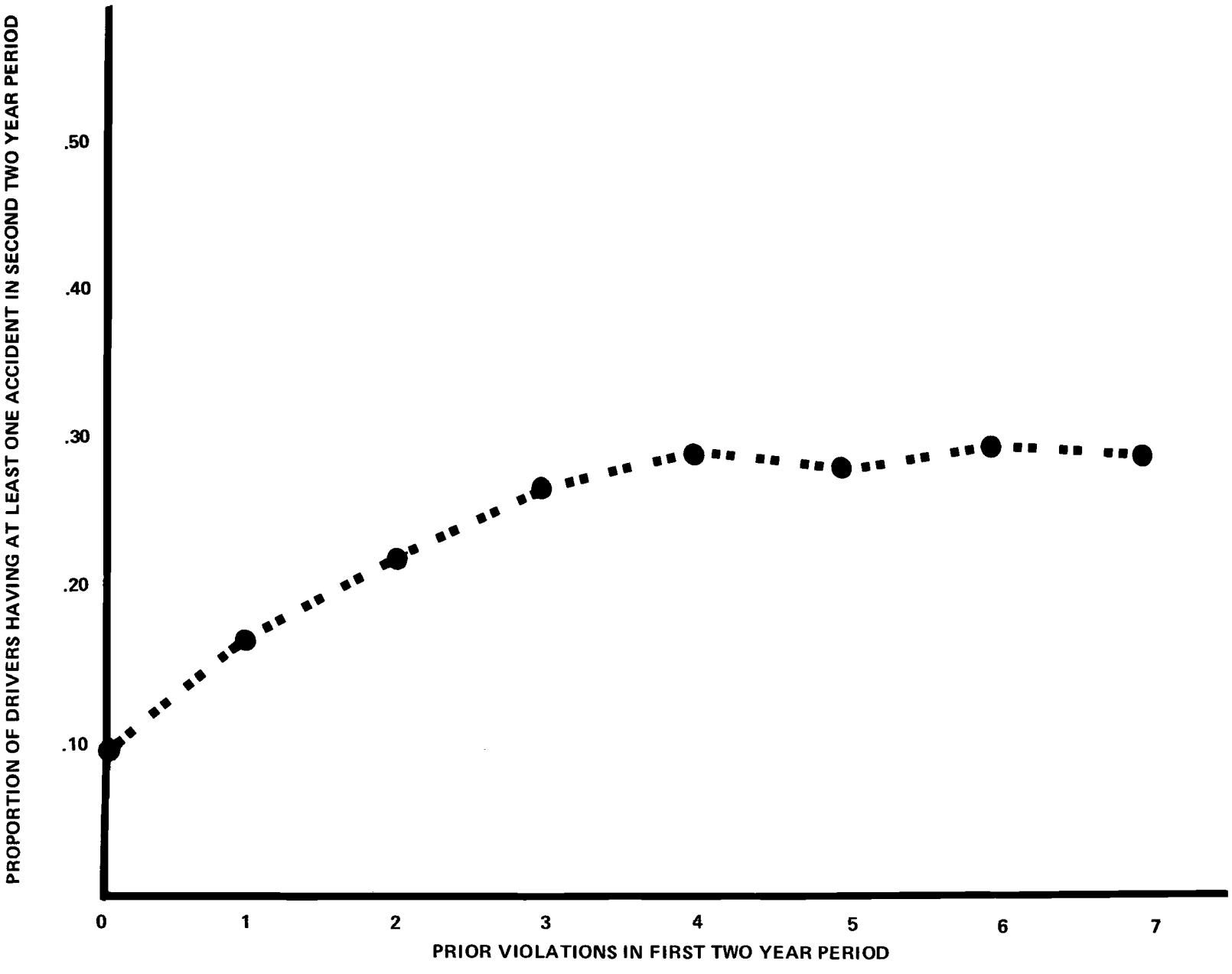

Figure 1. Proportion of drivers having at least one accident in a two-year period as a function of violations in a prior two-year period.

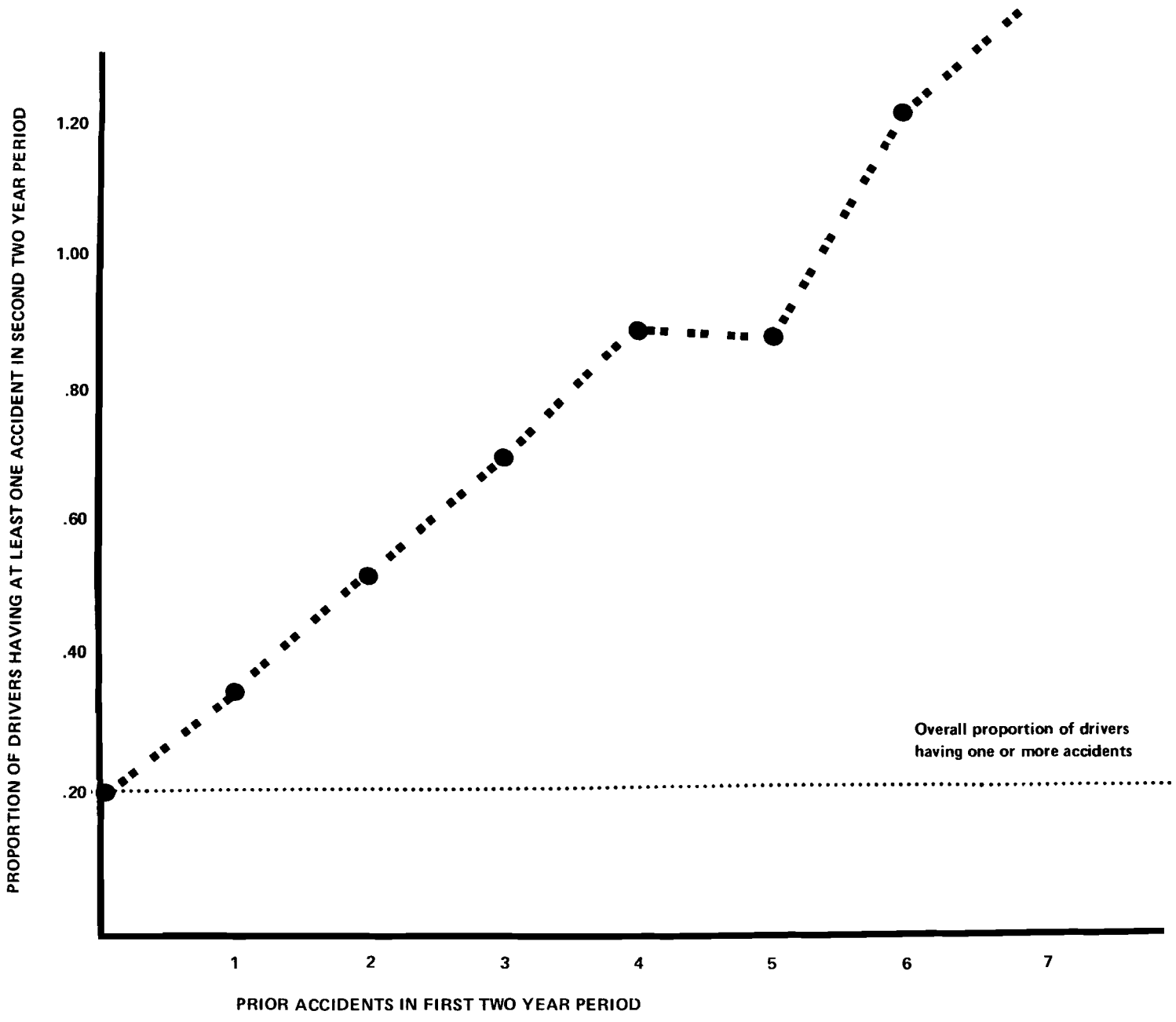

Figure 2. Proportion of drivers having at least one accident in a two-year period as a function of accidents in a prior two-year period.

However, Tables 3 and 3-A, which show violations in the second period as a function of violations in the first period, seem to indicate that this is *not* the case. For instance, the fifth row of 3-A shows that over 60 percent of the drivers having 4, 5, 6, and 7+ violations in the first period had one or more violations in the second period.

Thus accidents predict accidents and violations predict violations better than violations predict accidents.

The last rows of Tables 1-A and 2-A show average numbers of accidents as functions of prior violations and prior accidents, respectively. These are also shown in Figures 3 and 4 and exhibit the same sort of behavior as do proportions with one or more accidents (i.e., steadily increasing as a function of prior accidents and leveling off as a function of prior violations.)

It would, therefore, seem that from the standpoint of predicting relatively high likelihood of involvement in future accidents, past accident records would serve as a better predictor than would past violation records, though the ability to predict in either case is poor.

A more detailed analysis involving the fitting of certain types of curves to average numbers of accidents and proportions having one or more accidents is given in the appendix.

Table 3. Violation distributions over a two-year period as a function of violations in a prior two-year period.

Violations—first two-year period	Violations—second two-year period								Row Total
	0	1	2	3	4	5	6	7+	
0	1,787,466	243,548	48,882	11,848	3,438	1,124	394	235	2,096,935
1	206,451	62,580	20,052	6,329	2,106	729	261	137	298,645
2	40,193	19,077	8,462	3,386	1,308	465	208	117	73,216
3	9,930	6,090	3,261	1,462	683	298	113	70	21,907
4	2,854	1,874	1,263	643	331	145	62	52	7,224
5	888	679	482	281	133	74	33	27	2,597
6	386	234	175	124	61	33	16	13	1,042
7+	247	147	105	67	48	20	19	21	674
COLUMN TOTAL	2,048,415	334,229	82,682	24,140	8,108	2,888	1,106	672	2,502,240

G = .5046 Sd = .0012

ρ = .2138

τ = .0612 Sd = .0002

Table 3-A. Summary measures derived from Table 3.

No. of Violations During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of violation during first period	2,096,935	298,645	73,216	21,907	7,224	2,597	1,042	674	2,502,240
Proportion of total no. of drivers	.8380	.1194	.0293	.0088	.0029	.0010	.0004	.0003	*
No. of violations in second period by drivers having this no. of violations	400,237	136,265	55,783	22,388	9,114	3,775	1,552	1,111	630,225
Proportion of total violations in second period by drivers having this no. of violations in first period	.6351	.2162	.0885	.0355	.0145	.0060	.0025	.0018	*
Proportion of drivers with this no. of violations having at least one violation during second period	.1476	.3087	.4510	.5467	.6049	.6581	.6296	.6335	.1814
No. of drivers having at least this no. of violations in first period	2,502,240	405,305	106,660	33,444	11,537	4,313	1,716	674	*
No. of violations in second period by drivers having at least this no. of violations	630,225	229,988	93,723	37,940	15,552	6,438	2,663	1111	*
Proportion of all drivers having at least this no. of violations	1.00	.1620	.0426	.0134	.0046	.0017	.0007	.0003	*
Proportion of all violations by drivers having at least this no. of violations	1.00	.3649	.1487	.0602	.0247	.0102	.0042	.0018	*
Average no. of violations in second period by drivers having this no. of violations	.1909	.4563	.7619	1.0220	1.2616	1.4536	1.4894	1.6484	.2519

*No additional information provided by this cell

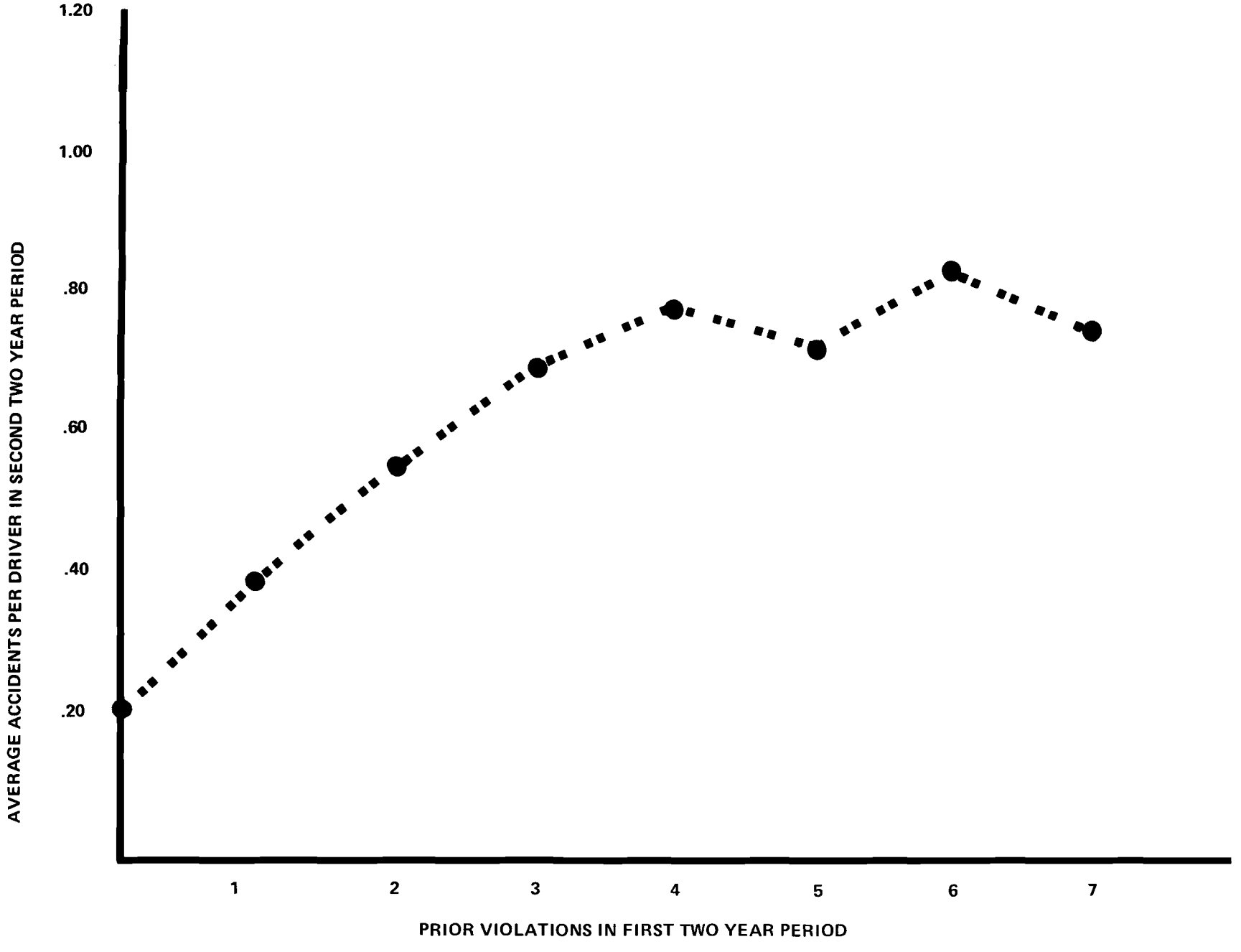

Figure 3. Average accidents per driver over a two-year period as a function of violations in a prior two-year period.

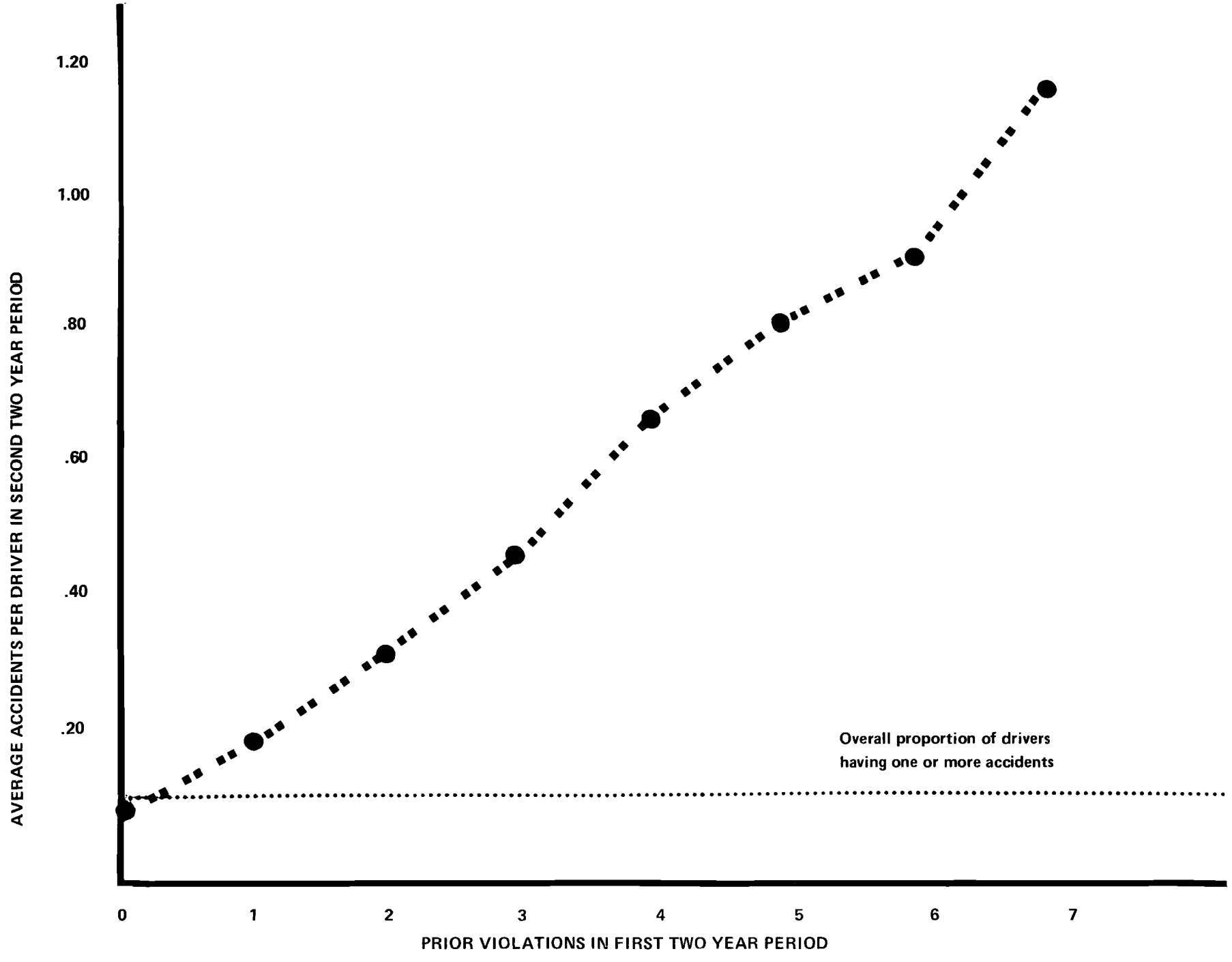

Figure 4. Average accidents per driver over a two-year period as a function of accidents in a prior two-year period.

Differences Due to Age of Driver

Results in the previous section dealt with all age groups. In order to investigate any differential effects associated with driver age, the data were divided into four driver age groups as of December 1, 1970. Tables were constructed for each of these age groups. Tables 4 through 7 show the results for the case of accidents versus prior accidents. These tables together with their accompanying summary tables (4A-7A) show, in general, very few differences among the various age groups. The most noticeable difference seems to be that the youngest age group, 22-25, is even less predictable than the other age groups. This can be seen from the correlations and also from certain of the summary measures such as average number of second period accidents, which starts at a higher level (.1537 for the 22 through 25-year-old group versus .1168 for the overall) for drivers having no accidents in the first two-year period, and then increases at a slower rate as a function of increasing first period accidents. Figure 5 shows average accidents per driver for the 22 through 25-year-old age group together with that for all age groups combined.

Table 4. Accident distributions for drivers aged 22-25 over a two-year period as a function of accidents in a prior two-year period.

Accidents—first two-year period—22-25 yr. old age grp.	Accidents—second two-year period—22-25 year old age group								Row Total
	0	1	2	3	4	5	6	7+	
0	276,081	36,196	4,967	715	105	28	7	2	318,101
1	33,615	7,164	1,358	237	36	5	0	0	42,415
2	4,639	1,339	313	63	18	4	1	0	6,377
3	648	239	54	17	3	2	0	1	964
4	73	41	18	3	2	0	0	0	137
5	14	7	0	2	0	0	1	0	24
6	2	2	1	0	0	0	0	0	5
7+	1	1	0	0	0	0	0	0	2
COLUMN TOTAL	315,673	44,989	6,711	1,037	164	39	9	3	368,075

G = .2913 Sd = .0052

ρ = .0877

τ = .0212 Sd = .0005

Table 4-A. Summary measures derived from Table 4.

No. of Accidents During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of accidents during first period	318,101	42,415	6,377	964	137	24	5	2	368,025
No. of drivers having this no. of accidents during first period	.8643	.1153	.0730	.0026	.0004	.0001	.0000	.0001	*
Proportion of total no. of drivers	48,891	10,760	2,252	427	94	19	4	1	62,448
No. of accidents in second period by drivers having this no. of accidents	.7829	.1723	.0361	.0008	.0015	.0003	.0001	.0000	*
Proportion of drivers with this no. of accidents having at least one accident	.1321	.2075	.2725	.3278	.4672	.4167	.6000	.5000	.1439
No. of drivers having at least this no. of accidents in first period	368,025	49,924	7,509	1132	168	31	7	2	*
No. of accidents in second period by drivers having at least this no. of accidents	62,448	13,557	2,797	545	118	24	5	1	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.1357	.0204	.0031	.0005	.0001	.0000	.0001	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.2171	.0448	.0087	.0019	.0004	.0001	.0000	*
Average no. of accidents in second period by drivers having this no. of accidents	.1537	.2537	.3531	.4429	.6861	.7917	.8000	.5000	.1697

*No additional information provided by this cell

Table 5. Accident distirbutions for drivers aged 26-39 over a two-year period as a function of accidents in a prior two-year period.

Accidents—first two-year period—26-39 yr. old age group	Accidents—second two-year period—26-39 yr. old age group								Row Total
	0	1	2	3	4	5	6	7+	
0	709,649	74,774	8,311	1,017	168	30	6	1	783,956
1	68,827	13,273	2,291	398	71	11	1	1	84,873
2	7,817	2,286	576	128	28	6	2	0	10,843
3	1,064	434	115	38	11	2	1	0	1,667
4	149	75	28	6	2	1	1	0	262
5	28	10	6	3	1	0	1	1	50
6	3	3	1	0	0	0	0	0	7
7+	0	0	0	0	0	0	0	0	0
COLUMN TOTAL	787,537	90,855	11,328	1,590	281	50	12	5	891,658

G = .3601 Sd = .0036

ρ = .0969

τ = .0195 Sd = .0003

Table 5-A. Summary measures derived from Table 5.

No. of Accidents During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of accidents during first period	793,956	84,873	10,843	1,667	262	50	7	0	891,658
Proportion of total no. of drivers	.8904	.0952	.0122	.0019	.0003	.0001	.0001	.0000	*
No. of accidents in second period by drivers having this no. of accidents	95,312	19,401	3,976	852	168	48	5	0	119,762
Proportion of total accidents by drivers having this no. of accidents	.7958	.1620	.0332	.0071	.0014	.0004	.0000	.0000	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.1062	.1891	.2791	.3617	.4313	.4400	.5714	.0000	.1168
No. of drivers having at least this no. of accidents in frist period	891,658	97,702	12,829	1,986	319	57	7	0	*
No. of accidents in second period by drivers having at least this no. of accidents	119,762	24,450	5,049	1,073	221	53	5	0	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.1096	.0144	.0022	.0004	.0001	.0001	.0000	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.2042	.0422	.0090	.0018	.0004	.0000	.0000	*
Average no. of accidents in second period by drivers having this no. of accidents	.1200	.2286	.3667	.5111	.6412	.9600	.7143	.0000	.1343

*No additional information provided by this cell

Table 6. Accident distributions for drivers aged 40-59 over a two-year period as a function of accidents in a prior two-year period

Accidents—first two-year period—40-59 yr. old age group	Accidents—second two-year period—40-59 yr. old age group								Row Total
	0	1	2	3	4	5	6	7+	
0	762,592	72,080	6,604	679	87	11	3	3	842,059
1	66,875	11,048	1,644	261	43	9	1	0	79,881
2	5,928	1,540	365	80	13	5	1	0	7,932
3	629	265	85	19	6	0	0	1	1,005
4	76	35	20	3	3	0	1	0	138
5	10	9	0	1	0	0	0	0	20
6	1	1	1	0	0	0	0	0	3
7+	2	2	2	0	0	0	0	0	6
COLUMN TOTAL	836,113	84,980	8,721	1,043	152	25	6	4	931,044

G = .3367 Sd = .0041

ρ = .0793

τ = .0141 Sd = .0002

Table 6-A. Summary measures derived from Table 6.

No. of Accidents During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of accidents during first period	842,059	79,881	7,932	1,005	138	20	3	6	931,044
Proportion of total no. of drivers	.9044	.0858	.0085	.0011	.0001	.0000	.0000	.0001	*
No. of accidents in second period by drivers having this no. of accidents	87,767	15,342	2,593	523	102	12	3	6	106,348
Proportion of total accidents by drivers having this no. of accidents	.8253	.1443	.0244	.0049	.0010	.0001	.0000	.0001	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.0944	.1628	.2526	.3741	.4493	.5000	.6667	.6667	.1020
No. of drivers having at least this no. of accidents in first period	931,044	88,985	9,104	1,172	167	29	9	6	*
No. of accidents in second period by drivers having at least this no. of accidents	106,348	18,581	3,239	646	123	21	9	6	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.0956	.0098	.0013	.0002	.0000	.0000	.0000	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.1747	.0305	.0061	.0012	.0002	.0001	.0001	*
Average no. of accidents in second period by drivers having this no. of accidents	.1042	.1921	.3269	.5204	.7391	.6000	1.00	1.00	.1142

*No additional information provided by this cell

Table 7. Accident distributions for drivers aged 60 and above over a two-year period as a function of accidents in a prior two-year period.

Accidents—first two-year period—60+ yr. old age group	Accidents—second two-year period—60+ yr. old age group								Row Total
	0	1	2	3	4	5	6	7+	
0	254,255	23,728	2,198	228	46	2	3	1	280,461
1	23,367	3,878	549	90	22	4	1	0	27,911
2	2,083	529	115	29	6	3	1	1	2,767
3	205	70	30	7	4	1	0	0	317
4	23	11	7	2	4	0	0	0	47
5	3	0	1	1	0	0	0	0	5
6	1	0	1	1	0	0	0	0	3
7+	0	1	0	0	1	0	0	0	2
TOTAL	279,937	28,217	2,901	358	83	10	5	2	311,513

G = .3385 Sd = .0070

ρ = .0810

τ = .0147 Sd = .0004

Table 7-A. Summary measures derived from Table 7.

No. of Accidents During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of accidents during first period	280,461	27,911	2,767	317	47	5	3	2	311,513
Proportion of total no. of drivers	.9003	.0896	.0089	.0010	.0002	.0000	.0000	.0000	*
No. of accidents in second period by drivers having this no. of accidents	29,027	5,360	898	172	47	5	5	5	35,519
Proportion of total accidents by drivers having this no. of accidents	.8172	.1509	.0253	.0048	.0013	.0001	.0001	.0001	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.0934	.1628	.2472	.3533	.5106	.4000	.6667	1.00	.1014
No. of drivers having at least this no. of accidents in first period	311,513	31,052	3,141	374	57	10	5	2	*
No. of accidents in second period by drivers having at least this no. of accidents	35,519	6,492	1,132	234	62	15	10	5	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.0997	.0101	.0012	.0002	.0000	.0000	.0001	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.1828	.0319	.0066	.0017	.0004	.0003	.0001	*
Average no. of accidents in second period by drivers having this no. of accidents	.1035	.1920	.3245	.5426	1.00	1.00	1.6667	2.500	.1140

*No additional information provided by this cell

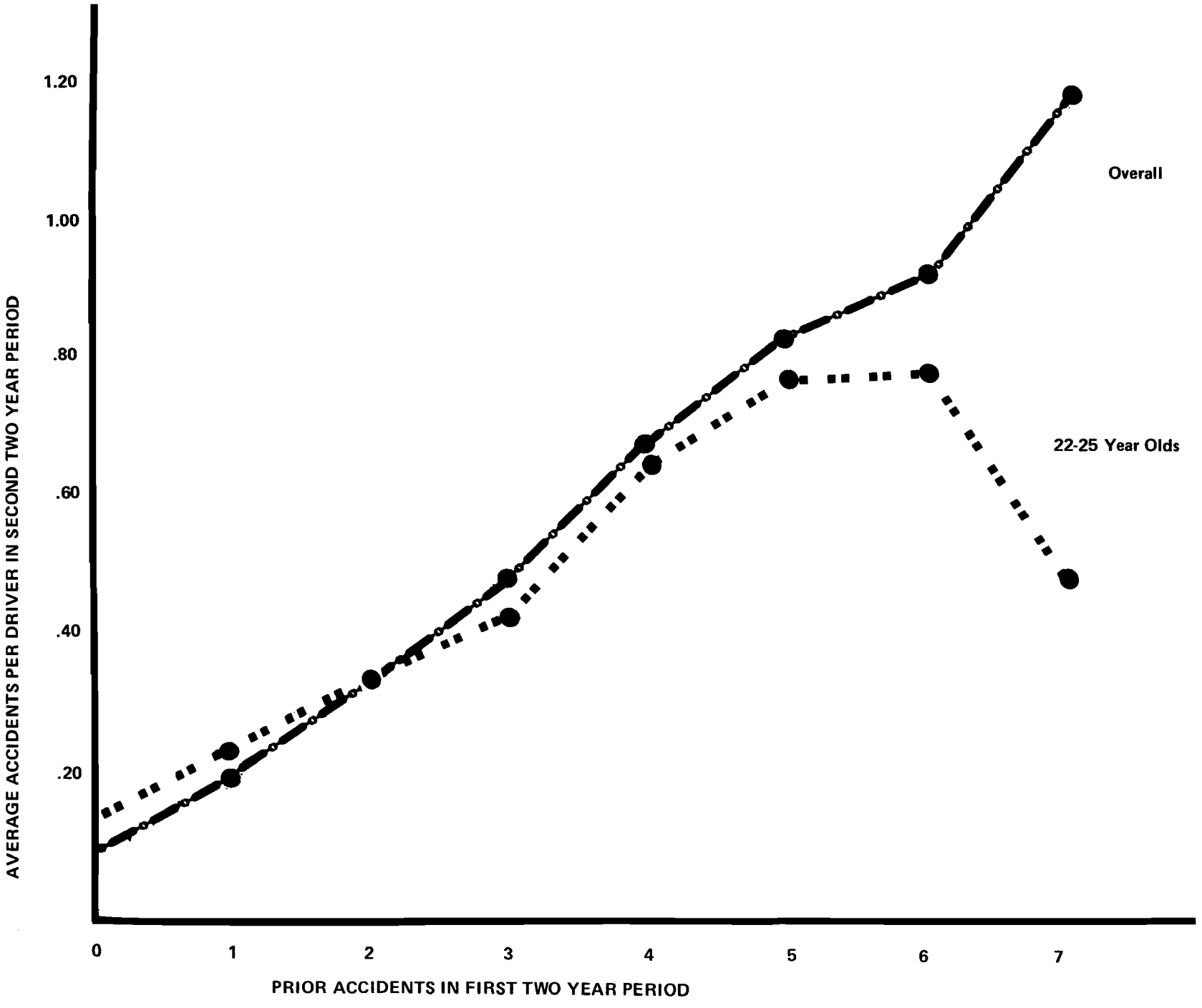

Figure 5. Comparison of average accidents per driver as a function of prior accidents for 22 through 25-year-old drivers with all age groups.

Effects of Relative Lengths of Time Periods

While most of analyses involved comparisons of driving performance during a two-year period with that in the preceding two-year period, certain analyses were done by dividing the four-year observation period in a different manner. Table 8 shows accidents in a one-year period as a function of accidents in the preceding three years while Table 9 shows accidents in a three-year period as a function of accidents in the preceding one-year period. Comparing these tables with Table 2 it can be seen that the correlations increase monotonically with the length of the second time period. This would seem to be due to two factors. First, the shorter prior time periods give rise to more coarsely defined groups, (i.e., K accidents in a one-year period is a worse record than K accidents in a two year period for any integer K = 0, 1, 2, . . . but not as bad as 2K accidents in a two-period). Secondly, the longer second observation periods provide more opportunity for the observation of the relatively rare occurrence of accidents. The effects of these two factors can also be seen in Figure 6 which shows average number of accidents per driver as a function of prior accidents for the three situations. The first factor gives rise to the different slopes of the curves while the second factor gives rise to the differences in intercepts.

The effects of the relative lengths of the two time periods will be discussed further in the next section.

Table 8. Accident distributions over a one-year period as a function of accidents in a prior three-year period.

Accidents—first three-year period	Accidents—second one-year period						Row Total
	0	1	2	3	4	5	
0	2,002,577	104,048	5,931	438	30	5	2,113,029
1	295,414	26,776	2,362	231	16	9	324,808
2	45,203	6,255	811	102	12	3	52,386
3	7,666	1,577	247	34	2	2	9,528
4	1,441	375	80	11	3	0	1,910
5	300	83	30	10	1	0	424
6	82	20	13	0	0	0	115
7+	25	4	7	3	1	0	40
COLUMN TOTAL	2,352,708	139,138	9,481	829	65	19	2,502,240

G = .3360 Sd = .0026

ρ = .0762

τ = .0131 Sd = .0001

Table 8-A. Summary measures derived from Table 8.

No. of Accidents During First Time Period	0	1	2	3	4	5	6	7+	0+
No. of drivers having this no. of accidents during first period	2,113,029	324,808	52,386	9,528	1,910	424	115	40	2,502,240
Proportion of total no. of drivers	.8445	.1298	.0209	.0038	.0007	.0002	.0001	.0000	*
No. of accidents in second period by drivers having this no. of accidents	117,369	32,302	8,246	2,191	580	177	46	31	160,942
Proportion of total accidents by drivers having this no. of accidents	.7293	.2007	.0512	.0136	.0036	.0011	.0003	.0002	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.0523	.0905	.1371	.1954	.2455	.2925	.2870	.3750	.0598
No. of drivers having at least this no. of accidents in first period	2,502,240	389,211	64,403	12,017	2,489	579	155	40	*
No. of accidents in second period by drivers having at least this no. of accidents	160,942	43,573	11,271	3,025	834	254	77	31	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.1555	.0257	.0048	.0010	.0002	.0001	.0000	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.2707	.0700	.0188	.0052	.0016	.0005	.0002	*
Average no. of accidents in second period by drivers having this no. of accidents	.0555	.0994	.1574	.2300	.3037	.4175	.4000	.7750	.0643

*No additional information provided by this cell

Table 9. Accident distributions over a three-year period as a function of accidents in a prior one-year period.

Accidents—first one-year period	Accidents—second three-year period								Row Total
	0	1	2	3	4	5	6	7+	
0	2,002,577	303,379	45,731	7,711	1,544	334	81	24	2,361,381
1	96,083	26,662	6,217	1,490	367	91	24	14	130,949
2	5,516	2,407	821	226	79	25	10	7	9,091
3	386	206	91	34	10	4	2	1	734
4	29	21	14	4	5	1	0	0	74
5	6	1	2	1	0	0	0	1	11
COLUMN TOTAL	2,104,607	332,677	52,876	7,466	2,005	455	117	47	2,502,240

G = .3564 Sd = .0020

ρ = .0807

τ = .0136 Sd = .0001

Table 9-A. Summary measures derived from Table 9.

No. of Accidents During First Time Period	0	1	2	3	4	5	0+
No. of drivers having this no. of accidents during first period	2,361,381	130,949	9,091	734	74	11	2,502,240
Proportion of total no. of drivers	.9437	.0523	.0036	.0003	.0000	.0000	*
No. of accidents in second period by drivers having this no. of accidents	426,474	45,732	5,277	569	86	15	478,153
Proportion of total accidents by drivers having this no. of accidents	.8919	.0956	.0110	.0012	.0002	.0000	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.1519	.2663	.3932	.4741	.6081	.4545	.1589
No. of drivers having at least this no. of accidents in first period	2,502,240	140,859	9,910	819	85	11	*
No. of accidents in second period by drivers having at least this no. of accidents	478,153	51,679	5,947	670	101	15	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.0563	.0040	.0003	.0000	.0000	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.1081	.0124	.0014	.0002	.0000	*
Average no. of accidents in second period by drivers having this no. of accidents	.1806	.3492	.5805	.7752	1.1622	1.3636	.1911

*No additional information provided by this cell

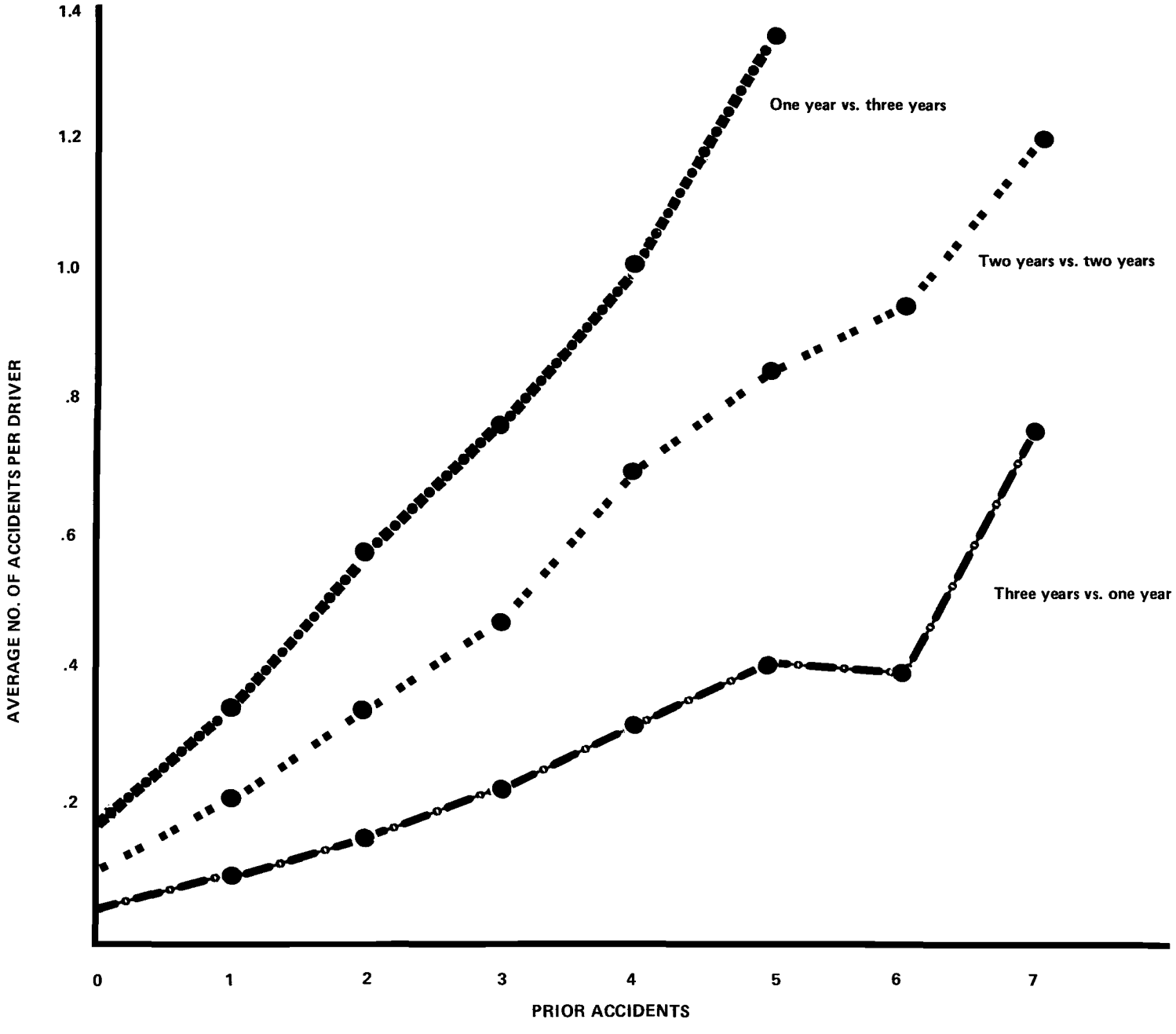

Figure 6. Effects of observation periods of varying lengths on average accidents per driver as a function of prior accidents.

CONCLUSIONS

From the contingency tables and their accompanying summary tables, several conclusions concerning the predictability of accident involvement can be drawn.

The previous violation record of a driver is not a good predictor of his future accident involvement. Thus, for instance, Table 1-A shows that of the 324,077 accidents occurring during the last two years of the study period 229,831 or 70.92 percent of them involved drivers who had *no* violations during the preceding two year period. Thus, 71 percent of *all* of accidents in the second time period were accounted for by people with *no* recorded offenses the previous two years—the very people who would be assumed to have a clean record later.

Now let us look at the drivers who had the *worst* earlier records. Drivers with four or more violations in the first two-year period (.46 percent of all drivers) went on to become involved in 4,561 accidents in the second two-year period. These 4,561 accidents represent 1.4 percent of the total accidents that later occurred. Moreover, about 70 percent of the drivers having *four or more violations* in the *first* two years had *no accidents* in the *second* two years. This 70 percent figure seems to hold constant with increasing numbers of violations in the first time period, (i.e., even for drivers with seven or more violations 70.33 percent had no accidents in the second period). This together with Table 3-A, which shows that 63.35 percent of the drivers with 7 or more violations in the first two years *also had violations* in the following two-year period, seems to indicate that there are drivers who continue to accumulate violations, but who are not involved in accidents.

The previous accident record, on the other hand, seems to be a slightly better predictor of accidents in the sense of identifying groups of drivers with relatively high likelihoods of being involved in future accidents. Thus, the fifth row of Table 2-A shows that 35.59 percent of all drivers having three accidents in the first two year period had at least one accident in the second period, 45.03 percent of those having four accidents in the first period had one or more in the second period, etc. Figure 7 shows the proportion of drivers having at least one accident in second two-year period having had k or more accidents in the first two-year period, as a function of k. From this figure it can be seen that 37.14 percent of those drivers who had three or more accidents during the first time period were involved in accidents in the second, 45.71 percent who had four or more in the first period had accidents in the second, and so on up to 70 percent for those who had seven or more accidents in the first period. It should be noted, however, that as the *probability* of future accident involvement gets higher the *number of drivers concerned and the number of accidents they have* in the second time period diminish to a very small proportion of the respective totals. Thus, from rows 6, 7, 8, and 9 of Table 2-A, it can be seen that 4,664 drivers (0.19 percent of the total) had three or more accidents in the first two years, and these drivers had 2,498 accidents (0.77 percent of all accidents) in the second two years; 711 drivers had four or more accidents in the first period (0.03 percent), and they accounted for 524 accidents in the second period (0.16 percent); etc. Again, as in the case of prior violations, the vast majority of accidents (80.54 percent) in the second time period involved drivers who showed *no* accident symptoms during the first two year period.

In summary, then, previous accident records seem to be somewhat better discriminators of relatively high risk driver groups than are previous violation records. The numbers of drivers in the groups so identified are quite small and the accidents these drivers are involved in make up only a small proportion of the total. Also, many who would be predicted to have a future accident do *not,* while most future accidents are experienced by drivers who would *not* be predicted to have them.

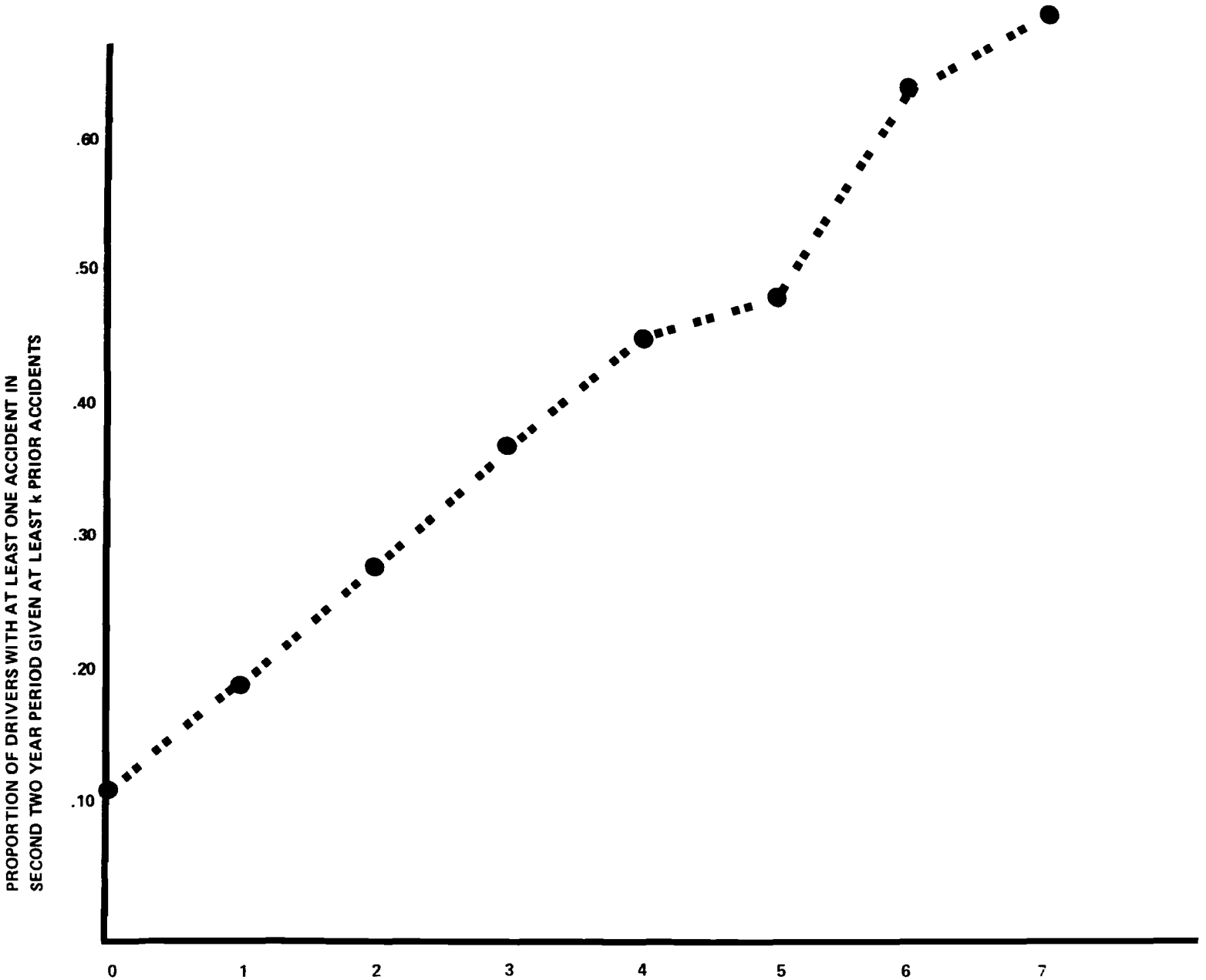

Figure 7. Proportion of drivers having at least one accident in a two-year period as a function of having no fewer than a given number of accidents in a prior two-year period.

Age Effects

Very few differences can be noted with respect to the predictability of driving performance from post records as a function of driver age. The one exception to this is the generally weaker relationships between the driving records over the two consecutive time intervals for the youngest (22 through 25 year old) group as discussed in the analysis section. This same type of relationship is also observed between violations and accidents, and violations and violations over the two time periods for this youngest age group.

Tables 10 and 10-A show the relationships between accident records over two consecutive one-year periods for drivers who were twenty-one years of age during the last year of the study period. Here, an even greater lack of correlation is seen, but the shorter time intervals certainly would contribute to this effect. In general, it would seem that for young drivers, previous driving records may be an even less valid indicator of future accident involvement than for older drivers.

Effects of Relative Length of Time Periods

The observed effects on the relationships between accident records in two adjacent time periods produced by varying the relative lengths of the time periods (i.e., 1 year vs 3 years, 2 vs 2, 3 vs 1) were discussed in the analysis section. For the purpose of identifying high risk groups of drivers it might be more informative to keep the length of one of the time intervals fixed and vary the other, particularly if some criteria could be specified in advance. For example, if one wished to identify a group of drivers for whom it was estimated that at least one half would be involved in one or more accidents in the next two years, one might select those drivers having four or more accidents in the last two years, five or more in the last three years, etc., or some combination of these. By then comparing the numbers of drivers and accidents involved, it might be possible, in some sense, to determine an optimal record length for this purpose. No attempt to do this sort of thing has been done in this study.

Selection and Selection Errors

If the decision is to be made to take some action against certain "high risk" drivers as determined by their past driving record in order to prevent future accidents, then it is important to know consequences of this decision, such as: How many drivers would be selected as being "high risk"? How many accidents would these drivers be expected to have in some future time period?, and how many of these drivers would be expected to have no accidents in the future period? (i.e., falsely or erroneously predicted as "high risk").

Table 11 provides this type of information for North Carolina drivers during the four-year study period, where the criterion for being selected was the number of violations during the first two years, and the accident numbers represent accidents involving these drivers during the second two years. Table 12 provides the same information where the criterion for selection is based upon the number of accidents in the first two years.

Thus, the first row of Table 11 shows that if all drivers with one or more violations were selected out for some accident prevention activity, this would involve 405,305 drivers (16.20 percent of all North Carolina drivers). These drivers had 94,246 accidents during the following two years (29.08 percent of all accidents), which was in excess of the number they would be expected to have based on the overall average number of accidents per driver during this period. However, 327,709 of these drivers selected (80.85 percent) had *no* accidents during the second two-year

Table 10. Accident distributions of 21-year-old drivers over a one-year period as a function of accidents in a prior one-year period.

		Accidents—second one-year period						Row Total
		0	1	2	3	4	5	
Accidents—first one-year period	0	144,803	12,184	1,036	77	10	0	158,110
	1	13,118	1,767	232	26	1	1	15,145
	2	1,204	215	33	5	2	0	1,459
	3	113	23	7	1	0	0	144
	4	13	2	1	0	0	0	16
	5	0	0	0	1	0	0	1
COLUMN TOTAL		159,251	14,191	1,309	110	13	1	174,875

G = .2699 Sd = .0109

ρ = .0568

τ = .0095 Sd = .0005

Table 10-A. Summary measures derived from Table 10.

No. of Accidents During First Time Period	0	1	2	3	4	5	0+
No. of drivers having this no. of accidents during first period	158,110	15,145	1,459	144	16	1	174,875
Proportion of total no. of drivers	.9041	.0866	.0083	.0008	.0001	.0001	*
No. of accidents in second period by drivers having this no. of accidents	14,527	2,318	304	40	4	3	17,196
Proportion of total accidents by drivers having this no. of accidents	.8448	.1348	.0177	.0023	.0002	.0002	*
Proportion of drivers with this no. of accidents having at least one accident during second period	.0842	.1338	.1748	.2153	.1875	1.00	.0893
No. of drivers having at least this no. of accidents in first period	174,875	16,765	1,620	161	17	1	*
No. of accidents in second period by drivers having at least this no. of accidents	17,196	2,669	351	47	7	3	*
Proportion of all drivers having at least this no. of accidents in first period	1.00	.0959	.0093	.0009	.0001	.0001	*
Proportion of all accidents by drivers having at least this no. of accidents	1.00	.1552	.0204	.0027	.0004	.0002	*
Average no. of accidents in second period by drivers having this no. of accidents	.0919	.1531	.2084	.2778	.2500	3.00	.0983

*No additional information provided by this cell

Table 11. Characteristics of "High Risk" drivers selected by prior violations.

Criterion	No. of Drivers	% of Drivers	No. of Accidents	% of All Accidents	% of Excess*	No. With No Accidents	% of These Drivers
1 or more viol.	405,305	16.20	94,246	29.08	79.56	327,709	80.85
2 or more viol.	106,660	4.26	33,336	10.29	141.35	80,443	75.42
3 or more viol.	33,444	1.34	12,416	3.83	186.68	23,985	71.72
4 or more viol.	11,537	.46	4,561	1.41	205.28	8,124	70.42
5 or more viol.	4,313	.17	1,671	.52	199.18	3,051	70.74
6 or more viol.	1,716	.07	706	.22	217.70	1,203	70.10
7 or more viol.	674	.03	258	.08	195.59	474	70.33

* % Excess computed by

$$\% \text{ Excess} = \frac{(\text{No. of Acc.}) - (\text{No. of Drivers})(\text{avg. no. acc./drivers})}{(\text{No. of Drivers})(\text{avg. no. acc./drivers})}$$

Table 12. Characteristics of "High Risk" drivers selected by prior accidents.

Criterion	No. of Drivers	% of All Drivers	No. of Accidents	% of All Accidents	% of Excess	No. With No Accidents	% of These Drivers
1 or more Accidents	267,663	10.70	63,080	19.46	81.98	216,083	80.73
2 or more Accidents	32,583	1.30	12,217	3.77	189.54	23,399	71.81
3 or more Accidents	4,664	.19	2,498	.77	313.58	2,932	62.86
4 or more Accidents	711	.03	524	.16	469.10	386	54.29
5 or more Accidents	127	.005	113	.03	587.08	65	51.18
6 or more Accidents	28	.001	29	.009	699.78	10	37.04
7 or more Accidents	10	.0004	12	.004	826.64	3	30.00

period. Reading down the columns of Table 11, several things can be observed, as the number of violations permitted before selection increases. 1) The percent of excess (more than "their share" of accidents) increases to around 200 percent. 2) The numbers of drivers selected and accidents involving them decreases very rapidly. 3) The percent of selected drivers having no accidents (falsely selected) does not decrease below about 70 percent.

From Table 12 it can be seen that when the criterion for selection is the prior accident record, the percent of excess rises more with worsening record, and the percent of selected drivers having no accidents falls to 50 percent or lower. On the other hand, the number of drivers selected and preventable accidents becomes small even more rapidly.

Thus, in making such a decision one would have to choose between a criterion which would select out a relatively large number of drivers (a high proportion of whom would be expected to have no accidents) in order to have a chance to prevent a relatively large number of accidents or one which would select a lower proportion of drivers expected to have no accidents at the sacrifice of having a chance to prevent a much smaller number of accidents.

SUMMARY

The foregoing data demonstrate once again that automobile accidents are not a very predictable phenomenon. It perhaps will be worthwhile to summarize some implications of this finding.

First, the fact that accidents are not highly predictable does *not* mean that they are capricious events not subject to understanding or control. It *does* however indicate limitations in addressing the accident problem by seeking out a relatively small number of drivers thought to account for a large portion of the accident problem.

As the data indicate, the majority of accidents are sustained by the majority of drivers. This indicates that countermeasure programs that society supports must be manifold and widespread in nature. This is quite unfortunate because it means the resources that must be expended are greater.

If it *were* true that the bulk of the accident problem was caused by only a few drivers and that these drivers remain largely constant and were subject to identification, then the financial burden of countermeasure programs would be smaller. For example, if only a small percentage of the people consistently caused the bulk of accidents, the State could get along with a smaller highway patrol simply because it would be easier to keep this small group of drivers under surveillance. The fact is, however, that the highway patrol has to be large because it has to exercise a controlling influence over the great majority of drivers inducing "us" to drive safely and legally.

It is *because* the majority of the drivers have the majority of the accidents that we must have widespread driver licensing and driver education programs that reach all drivers rather than being able to focus on a small core of problem drivers. Similarly, roadways must provide safety everywhere at all times, just as vehicles must.

None of the foregoing is, however, intended to deny or underemphasize the fact that there is indeed a tiny percentage of drivers with bad records who tend to persist in their bad records in the next time period. Thus it can be demonstrated that the worst .5 percent of all drivers in North Carolina in one time period had 1.5 percent of the accidents in the second time period or three times their share of accidents in the *next* time period.

It certainly is the case that for any such group of drivers, no matter how small, it is worthwhile to have special programs if it can be shown that these drivers have several times their share of accidents. North Carolina and most other states have driver improvement techniques, medical review procedures, etc., for this reason.

Since, however, these drivers with three times their share of accidents still only account for a very small proportion of total accidents in the State, it is evident that such special programs cannot detract from emphasis on other more widespread programs. Nor can programs aimed at the consistently high risk driver be expected to achieve large statewide changes in the accident picture. Close review must be maintained on the relative cost of dealing with this tiny group of drivers versus the potential benefit even if the programs are successful.

Another factor is that even among so called "high risk" drivers a very significant proportion of them have no future accidents. This raises questions of fairness as to the kinds of programs. For example, if a very stringent suppressive program were brought to bear on drivers with a violation record in the previous time period, one has to face the consequences of the fact that the majority of this group are drivers who in fact would have clean accident records in the future.

This is somewhat analogous to a situation in which all persons found guilty are subjected to stringent treatment despite the fact that more than half were in fact innocent. This kind of ratio is generally unacceptable to people in the context of a punitive program.

It would be more socially acceptable however if positive programs were brought to bear on these drivers. Instead of doing something *to* them, the State would do something *for* them. The high risk group might be given remedial instruction, skill training, instruction in emergency vehicle operating techniques, etc. Again, there is the question of the cost vs the practical benefit when dealing with such a small proportion of drivers.

Accident repeaters are a small but important part of the overall problem. They deserve attention through ingenious and cost-effective programs. It should be realized, however, that countermeasures directed at the violation and accident repeater cannot bring about large reductions in total accidents but can bring about small and important gains.

APPENDIX

In view of the relatively small numbers of drivers having more than two or three accidents or violations in the first two-year period, the relationships between accidents and previous accidents and violations was further investigated by fitting the proportion of drivers having accidents in the second two-year period as functions of prior violations and prior accidents by a weighted least squares procedure.

Quadratic functions fit quite well in both cases. In the case of *prior violations* the fitted model was

$$P(v) = .0982 + .0818v - .00801v^2, \text{ (i.e.) } P(v) = .0982 + .0818v - .00801v^2$$
$$\qquad\qquad (.0002)\ (.00079)\ (.00027)$$

where the standard errors of the coefficients are shown in parentheses. The X^2 due to error for this model was $X^2_{5\ \text{d.f.}} = 7.9178$, $p = .16$.

As a function of *prior accidents* the fitted equation was

$$P(a) = .1038 + .0739\,a + .00312\,a^2,$$
$$\qquad\quad (.0002)\ (.0014)\quad (.00086)$$

$X^2_{5\ \text{d.f.}}$ (due to error) $= 6.6265$, $p = .25$.

Thus, the proportion having at least one accident in the second period is fit as an increasing function of both a (prior accidents) and a^2, while as a function of prior violations the proportion increases with v but decreases with v^2. Figure 8 shows graphs of these two weighted least squares curves. Their behavior is very similar to that of the proportions plotted in Figure 1 and 2.

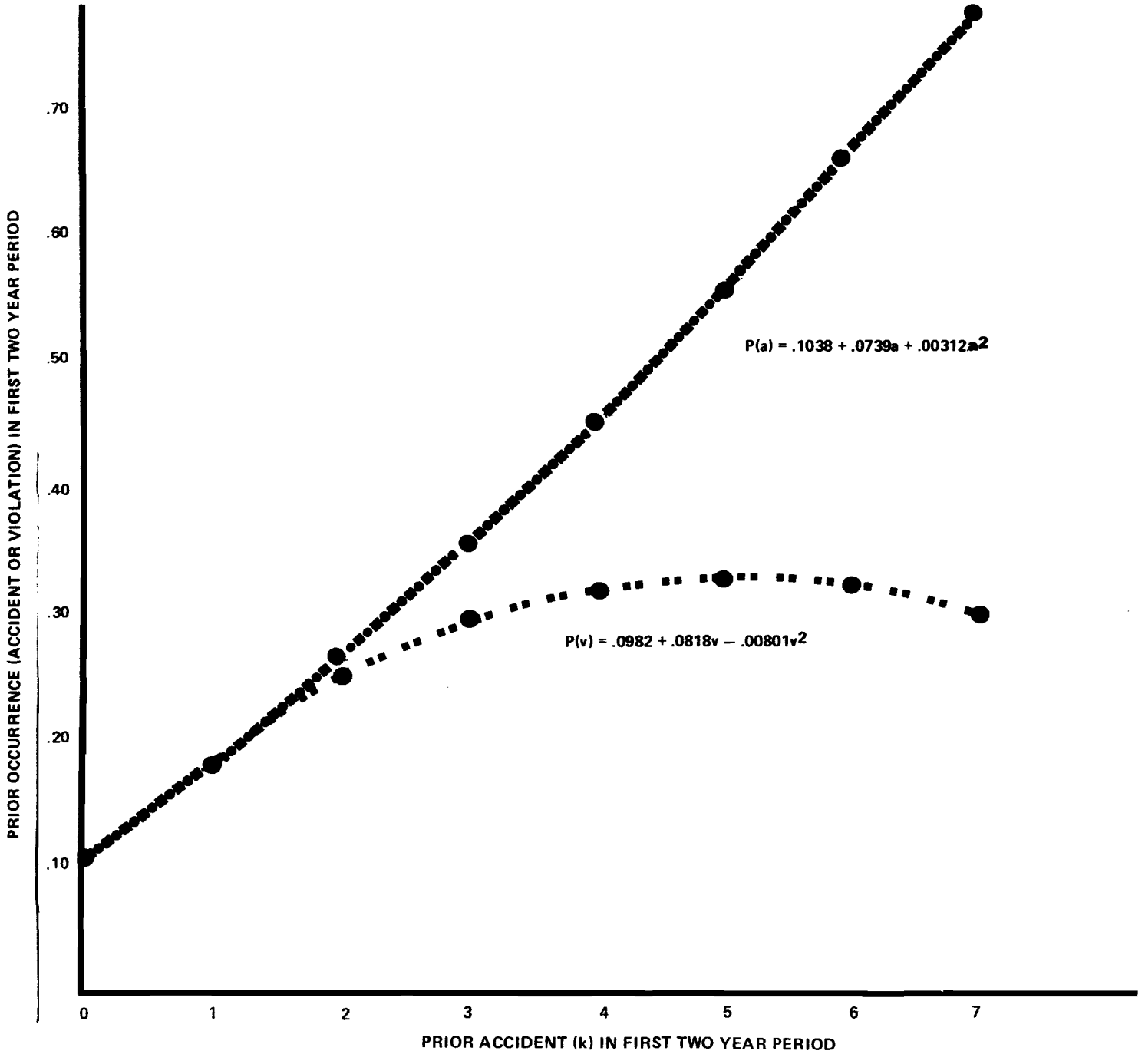

Figure 8. Weighted least squares fits to proportion of drivers having at least one accident as functions of prior violations and or prior accidents.