19670

14

The Analysis of State Driver Records: A Final Review

William L. Carlson
April 17, 1970

The Analysis of State Driver Records: A Final Review

The objectives of this paper are to present the results of our attempts to develop an improved driver prediction model and to present some interesting results from an extensive analysis of a sample of driver records.

This study is an extension of a previous study of Michigan driver records.[1] In the previous study 1071 driver records were analyzed and a mathematical model to predict accident involved drivers was developed. In addition the following conclusions were based upon the analysis:

1. The average number of accidents per driver for all drivers over the six and one-half year period (January 1, 1961 - June 1, 1967) is 0.42. Interestingly, a study conducted in California of 150,000 drivers during the years 1961 through 1963 reported an average number of accidents of 0.20 for the three year period.

2. Groups of problem drivers who have a significantly greater number of accidents than the average accident rate for the entire group of drivers can be identified. This identification has been accomplished with only those data currently available in official driver files.

3. The most significant single identifier of these problem drivers is the total number of motor vehicle offense convictions. Those drivers who did not incur any convictions over the six and one-half year period of this study had an accident rate which was almost one-third that of the average driver in the State of Michigan.

4. There are significant interactions among the total number of convictions and other factors as reported by the Michigan Driver Record-Keeping System. For example, of the drivers with convictions, (a) male drivers have a higher average number of accidents than female drivers, (b) drivers in the age group 20-25 have more accidents, and (c) chauffeurs or commercial drivers have a higher number of accidents than do persons having operator licenses. For the drivers who did not have any convictions, these differences did not occur.

5. This study should be extended to provide sharper tools for identifying groups of problem drivers. Two major types of expansion are needed.

    a. An increase in the number of factors considered. This would allow us to refine our identification of groups of problem drivers. For example, we are confident that the inclusion of additional alcohol-involvement data would significantly improve the predictive capability of the model.

    b. An increase in the number of drivers considered. This will allow a better cancelling of random factors not included in the study. It will also provide a means for both testing those hypotheses based on the present small sample and for generating new hypotheses.

This work is a direct result of conclusion number 5-a. Due to lack of funding it was not possible to obtain a large increase in the number of driver records. However, a new sample consisting of the records for 1331 drivers who were randomly sampled from a population of approximately 45,000 driver records obtained from the Michigan Department of State. These records represented only persons whose last names begin with the letter A. However, it is not believed that this introduced a serious bias.

In the analysis an attempt was made to determine whether improvement in the prediction of drivers likely to be involved in crashes (problem drivers) could be made by obtaining a large number of driver record variables. Thus we defined the driving experience of an individual as completely as possible using the existing driver record information. If this increase in number of variables, as compared to our previous study, is truly supplying new information then it should be possible to improve our capability to predict problem drivers.

As a result of an extensive analysis of this data, as reported below, it was determined that we could not improve the prediction of problem drivers relative to the previous study by using an in-

creased set of variables. There were two major reasons for this result:

1. Many of the variables occurred at only one level for most drivers. In particular, only a small percentage of drivers had serious convictions such as reckless driving.

2. Many of the variables were correlated. In particular persons with one type of conviction were more likely to have other types of convictions than were drivers without the first type of conviction. Thus no new information resulted from the increase in number of variables.

Thus we conclude that if we are to improve our prediction of problem drivers, we must obtain variables which describe different characteristics from those found in the driver records. However, this approach is also risky as shown by Levonian's extensive study of driver characteristics.[2] In that study data representing thirty-five predictor variables - roughly classified as four personal variables, 24 vision variables, and seven driving variables were obtained by personal interview for 7430 California drivers. In addition their number of crashes and convictions were obtained from their driving record. From this data a multiple regression model was computed to predict accidents and convictions, the resulting model contained only driving exposure, age, sex, and marital status. In addition, the model developed, using one half the data, was applied to the remainder of the data. This model predicted apriori forty-six percent of the drivers who were involved in crashes or driving convictions over a three year period.

In addition, this study indicates that the Michigan Point system predicts problem drivers just as accurately as does a problem driver prediction model developed from a previous analysis of driver record data. It should be added that both of these approaches have a great deal of room for improvement. However, at present it is more reasonable to continue with the present Michigan Point System.

Thus it is strongly recommended that driver record variables be used only as dependent variables indicating an individuals response within the system.

A number of very interesting results were obtained from the analysis of the data. The major conclusions are as follows:

Discussion

This project has been conducted as an extension of the driver record study discussed in the paper "Identifying the Problem Driver from State Driver Records." The approach was to obtain the complete computerized driver record for a sample of Michigan drivers and to perform an extensive data analysis of the variables in the driving record. A total of 238 variables were defined and obtained for a sample of 1,331 driver records. The variables obtained included:

1. Number of occurrences of all driving convictions

2. Number of occurrences of various degrees of accident severity

3. Number of occurrences of various types of action taken against the driver

4. Accidents and total convictions for 2 three year periods in addition to a 6 year period. This was done to determine whether or not accident involvement can be predicted from driving convictions occurring during a previous period.

A large number of univariate and bivariate analyses were performed. The major conclusions are as follows:

1. Persons with convictions are more likely to have accidents than are persons without convictions. (See Table 1)

2. Persons having accidents in the first 3 year period are more likely to have accidents in the second 3 year period. (See Table 2)

3. Persons having convictions in first 3 year period are more likely to have accidents in the second 3 year period. (See Table 3)

4. A 3 year period is long enough to show a significant relationship between convictions and accidents. (See Table 4)

5. Persons with DUIL convictions are more likely to have accidents than are all of the other drivers in the sample. (See Table 5)

6. Persons with Reckless Driving convictions are more likely to have accidents than are all of the other drivers in the sample. (See Table 6)

7. A large number of bivariate distributions were obtained for sub-groups of the total population. Some of these include:

   a. Relationship between convictions and accidents for drivers age 25 and younger and for drivers age 26 and older. The previously established relationships continue to hold in these sub-groups.

   b. Relationship between particular conviction types (e.g., DUIL, reckless driving, etc.) and accidents within the sub-group of drivers having any convictions. In general, there does not appear to be additional predictive power in adding the type of conviction to the information that a conviction has occurred. This appears due to the fact that there are not many occurrences of these particular conviction categories. In addition, we might suggest that since convictions are not a perfect predictor of accident involvement, special conviction types are predicting differences near the noise level.

8. Figure 1 presents the average number of accidents per driver for a number of sub-groups described by the indicated variables. This sub-division was generated by the AID algorithm. The average number of accidents per driver is .30 for the 6 year period as compared to .42 for a 6.5 year period as presented in our first paper. An analysis of the data by a 3 year period indicated an average of .10 accidents per driver during the period 3/62-3/65 and an average of .20 accidents per driver during the period 3/65-3/68. A possible explanation is that certain older less serious accidents were dropped from the driver record in the process of coding accidents for the computer. The following driver record variables were included as possible accident predictors:

   1. Number of DUIL convictions
   2. Number of reckless driving convictions
   3. Total number of convictions
   4. Number of convictions with Financial Responsibility Follow-up

5. Number of times failure to answer Summons
6. The sending of a Warning letter
7. A request for driver to appear for re-exam
8. Number of Felony with auto convictions*
9. Number of cycle or scooter convictions*
10. Number of speed related convictions
11. Number of convictions involving improper turn
12. Number of convictions involving minor moving violations
13. Number of convictions involving equipment violations
14. Number of convictions involving obstruction or load violations
15. Number of convictions involving license violations
16. An indication of miscellaneous actions against drivers such as cancelled license or denied license*
17. Driver age
18. Driver sex
19. Original or renewal license

The marked (*) items did not occur frequently enough to be considered as predictors, and a number of the items were at the borderline of the possible prediction level. The strongest predictor variable in this group was total number of driving convictions. This agrees with the results of our first study. It should also be noted that of the drivers with four or more driving convictions the younger drivers had more accidents. This result also agrees with our first study. Certain other sub-groups were established by combinations of speed related and minor moving violations.

9. The predictor model developed in the previous paper was applied to this sample of drivers for the entire 6 year period. In addition, the model was applied to the first 3 year period (3/62 - 3/65) and used to predict accident involvement in the second 3 year period (3/65 - 3/68). The conclusions are as follows:

    a. Table 7 indicates the result of applying the prediction model to the entire 6 year period. As shown, 38 out of 84 drivers reported as having 2 or more accidents were predicted as having 1 or more accidents. This level of predictability was not as good as that achieved in the original study. The predictability is very comparable

to that achieved through application of the Michigan point system, as indicated in the previous study. Therefore, it can be concluded that the predictor model does predict persons who are over-involved in accidents, but it does not do this job better than the present point system.

b. Table 8 indicates the result of applying the prediction model to the three year period (3/62 - 3/65) and using this score to predict accidents in the three year period (3/62 - 3/65). As shown, 35 out of 101 drivers reported as having one or more accidents were predicted as having one or more accidents.

c. Table 9 indicates the result of applying the prediction model to the first three year period (3/62 - 3/65) and using this score to predict accident involvement in the second three year period (3/65 - 3/68). As shown, 30 out of 210 drivers reported as having one or more accidents were predicted as having one or more accidents.

TABLE 1

Number of Drivers Involved in Accidents Having Driving Convictions for the Period 3/62-3/68

		Number of Accidents			
		0	1	2	3 or more
Number of Driving Convictions	0	598 (86.8%)	79 (11.5%)	10 (1.5%)	2 (.3%)
	1	180 (68.2%)	68 (25.8%)	11 (4.2%)	5 (1.9%)
	2	92 (70.8%)	26 (20.0%)	7 (5.4%)	5 (3.8%)
	3	30 (62.5%)	13 (27.1%)	5 (10.4%)	0
	4	17 (47.2%)	14 (38.9%)	3 (8.3%)	2 (5.6%)
	5 or more	46 (56.0%)	19 (23.2%)	12 (14.6%)	5 (6.1%)

TABLE 2

Number of Drivers Involved in Accidents During Period 3/62-3/65 Versus Involvement in Accidents During Period 3/65-3/68

		Number of Accidents 3/65 - 3/68			
		0	1	2	3
Number of Accidents 3/62-3/65	0	963 (83.8%)	154 (13.4%)	24 (2.1%)	8 (.6%)
	1	65 (75.6%)	16 (18.6%)	3 (3.5%)	2 (2.3%)
	2	9 (64.3%)	5 (35.7%)	0	0

TABLE 3

Number of Drivers Having Convictions in Period 3/62-3/65 Involved in Accidents During Period 3/65-3/68

		Accidents 3/65-3/68		
		0	1	2 or more
Convictions 3/62-3/65	0	819 (84.3%)	124 (12.8%)	28 (2.9%)
	1	130 (79.8%)	26 (16.0%)	7 (4.2%)
	2	43 (76.8%)	13 (23.2%)	0
	3 or more	45 (76.4%)	12 (20.4%)	2 (3.2%)

TABLE 4

Number of Drivers Having Convictions in Period 3/65-3/68
Involved in Accidents During Period 3/65-3/68

		Accidents 3/65 - 3/68		
		0	1	2 or more
Convictions 3/65-3/68	0	740 (89.5%)	79 (9.6%)	8 (.9%)
	1	179 (73.1%)	55 (22.4%)	11 (4.5%)
	2	71 (72.4%)	20 (20.4%)	7 (7.2%)
	3 or more	47 (59.5%)	21 (26.6%)	11 (13.9%)

TABLE 5

Number of Drivers Having Accidents in Period 3/62-3/68
Having DUIL Convictions

		Accidents 3/62 - 3/68		
		0	1	2 or more
DUIL Convictions	0	941 (77.3%)	212 (17.4%)	64 (5.3%)
	1	17 (68.0%)	5 (20.0%)	3 (12.0%)
	2	5 (71.5%)	2 (28.5%)	0

TABLE 6

Number of Drivers Having Accidents in Period 3/62-3/68
Having Reckless Driving Convictions

		Accidents 3/62 - 3/68		
		0	1	2 or more
Reckless Driving Convictions	0	940 (77.9%)	208 (17.2%)	59 (4.9%)
	1	20 (54.1%)	11 (29.7%)	6 (16.2%)
	2	3	2	0

TABLE 7

Application of Predictor Model to Six Year Driving Record

		Number of Persons Predicted		
		0 Accidents	1 or more Accidents	
Number of Persons Reported	0 & 1 Accident	918	228	1146
	2 or more Accidents	46	38	84
		964	266	1230

TABLE 8

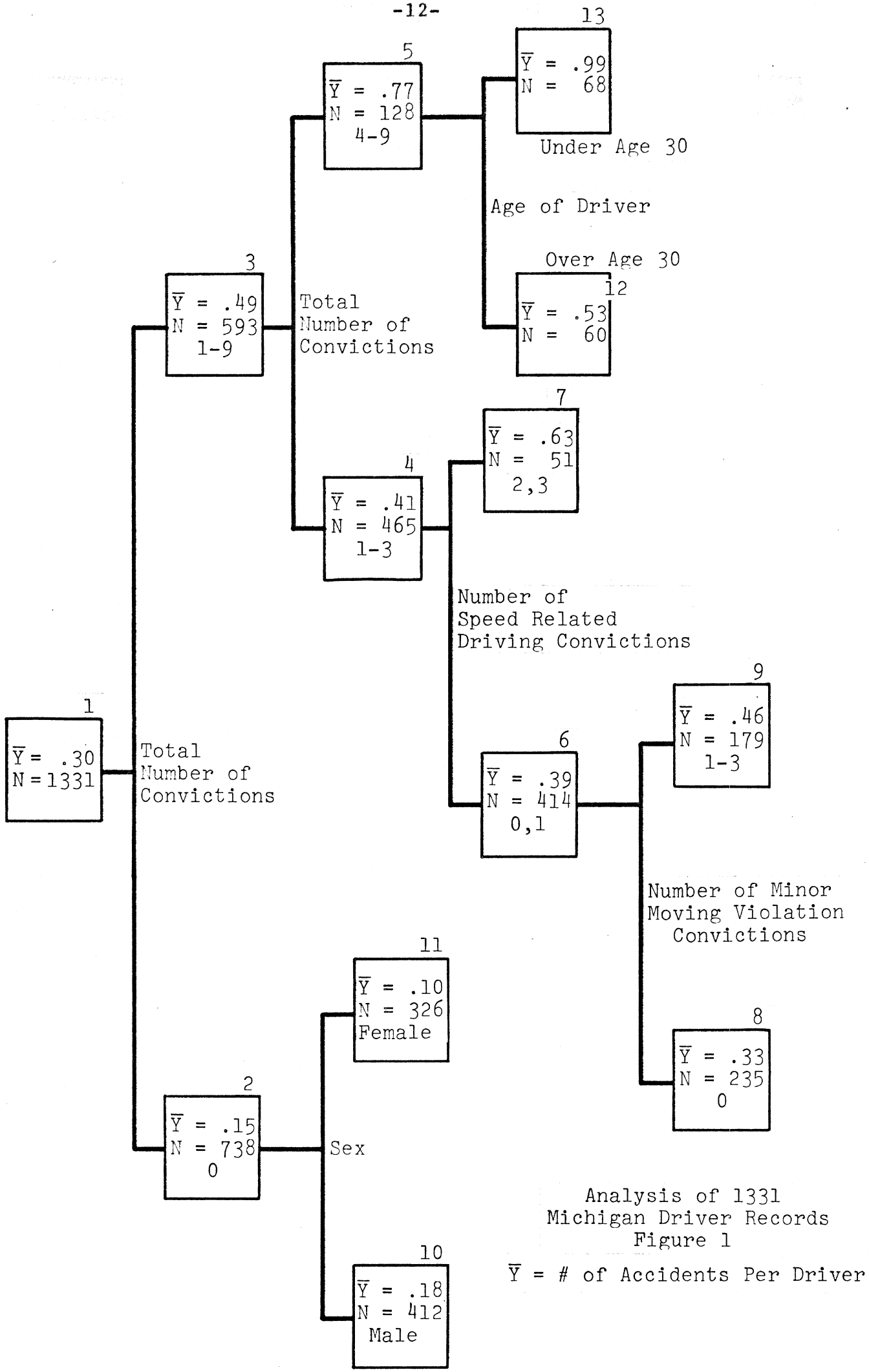
Application of Driver Predictor Model to Driving Records Covering the Period 3/62-3/65 and Predicting Accident Involvement for the Period 3/62-3/65

		Number of Persons Predicted		
		0	1 or more	
Number of Persons Reported	0	1084	101	1185
	1 or more	66	35	101
		1150	136	1286

TABLE 9

Application of Driver Predictor Model to Driving Records Covering the Period 3/62-3/65 and Predicting Accident Involvement for the Period 3/65-3/68

		Number of Persons Predicted		
		0	1 or more	
Number of Persons Reported	0	970	106	1076
	1 or more	180	30	210
		1150	136	



Analysis of 1331
Michigan Driver Records
Figure 1

$\bar{Y}$ = # of Accidents Per Driver

REFERENCES

1. Carlson, W.L., Identifying the Problem Driver from State Driver Records, Research No. 1, Highway Safety Research Institute, University of Michigan, Ann Arbor, May 1968.

2. Levonian, E., Case, H.W., Gregory R., Prediction of Recorded Accidents and Violations Using Non-Driving Predictors, Highway Research Record, No. 4, 1963 p. 50-61.