Driver Records and Crash Prediction

by Adrian Lund

From time to time, researchers on the staff of, or supported by, the Insurance Institute for Highway Safety develop analyses, findings, and observations that require publication in formats and on schedules more flexible and timely than those provided in journals. Research Notes is a vehicle provided by the Institute to make such material available on a timely basis. For more information concerning Research Notes or other Institute publications, write to Communications Department, Insurance Institute for Highway Safety, Watergate 600, Washington, DC 20037.

The Insurance Institute for Highway Safety is an independent, nonprofit scientific and educational organization dedicated to reducing the losses—deaths, injuries, and property damage—resulting from crashes on the nation’s highways. The Institute is supported by the American Insurance Highway Safety Association, the American Insurance Highway Safety Alliance, the National Association of Independent Insurers Safety Association and several individual insurance companies.

Copyright 1984
ISSN 0732-8532
Driver Records and Crash Prediction

Many states use driver licensing actions such as suspension or revocation to remove a minority of "problem drivers" with a history of crashes and violations from the highways. The basic premise for this approach is that the incidence of crashes can be reduced by identifying and restricting a small group of problem drivers who disproportionately contribute to the problem. Although it is possible that the existence of such programs also has a deterrent effect on other potential problem drivers, this is not the major rationale for such programs and is not the subject of this review.

To understand the contribution of problem drivers to the overall crash problem this Research Note reviews various studies that have addressed the question of how well past driving records can predict future crash involvement as shown on driver records. Given the reliance on driver records in these studies, it is important to recognize that the meaningfulness of such records varies considerably from state to state. A recent study (3) reported that only 81 percent of crashes that resulted in at least $500 in insurance claims appears in the driver records of the state (Wisconsin) with the most complete motor vehicle records; in 15 of the 17 states surveyed, fewer than 50 percent of such crashes appeared in the driver records. Thus, although driver records are better than self-reports, they are a weak source of crash information, and the estimates of past and future crash correlations may be somewhat conservative. Nevertheless, the results of studies of state driver records for California, Maryland, North Carolina, Texas, and Washington State were highly consistent and permit some conclusions to be drawn.* The basic findings were as follows:

- As a group, drivers with repeated traffic violations and/or crashes in a given period are several times more likely than drivers in the general population to have subsequent crashes. A study of the driving records of more than 100,000 California drivers during 1972-1974, for example, reported that drivers with two or more crashes in the first two years were three to six times more likely than drivers with no crashes to have a crash in the third year (3). Similar patterns were reported for drivers in North Carolina (12), as well as for Maryland and Washington State drivers involved in fatal crashes (3,4,11).

- Most drivers with violations and crashes on their records in a given period do not have recorded crashes during prior or subsequent periods. Although groups of high crash risk drivers can be identified on the basis of prior records, these drivers account for only a very small percentage of all crashes. In the three-year California study, 88 percent of those with two or more crashes in the first two years did not crash during the third year, whereas drivers with no crashes in the first two years accounted for 68 percent of the crashes in the third year (3). Similar results were reported in a four-year study of 2.5 million licensed drivers in North Carolina (12); drivers with no crashes in the first two years were least likely to have a crash in the second two years (figure 1a), but they accounted for over 80 percent of the crashes that did occur in the second two years (figure 1b). This apparent paradox occurs because, although drivers with no prior crashes are less likely to crash by a factor of two to seven, they outnumber "problem" drivers by a much larger factor. (They are 70 times more numerous than drivers with two or more crashes, as shown in figure 1c.) As a result, although there is a direct and replicable relationship between past crashes and future crashes on driver records, the level of association is quite low. This low association is illustrated by the very small amount of variation in subsequent crash records that can be predicted by variation in prior crash records (see column 5 of table, page 22).

- Traffic violation records predict future crash involvement slightly more accurately than crash involvement records. Still, records of violations predict only a small portion of all crashes (see column 6 of table, page 2). Even when age, sex, and other driver variables have been combined with records of violations and crashes, the resulting regression equations explain very little of the crash variation (7,8,9).

- Driving records predict involvement in severe crashes marginally better than they predict general crash involvement. Researchers in California found that drivers involved during a three-year period in nighttime crashes had more convictions than drivers involved in daytime crashes during the same period (7). Because nighttime crashes typically are more severe than daytime crashes, this finding supports the notion that problem drivers are more prevalent in severe crashes than in other crashes. However, other researchers have reported that the records of California drivers associated with fatal crashes during 1970-1971 had similar relative predictive weightings as

---

* It is important to remember, also, that any licensing action such as suspension or revocation is based on the same motor vehicle records.
Percent of Variation in Subsequent Crashes Accounted for by Past Driving Record

<table>
<thead>
<tr>
<th>State</th>
<th>Study (Ref. No.)</th>
<th>Pre-Crash Period (years)</th>
<th>Post-Crash Period (years)</th>
<th>Predictor: Past Crashes</th>
<th>Predictor: Past Violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>CA</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>&lt;1%</td>
<td>*</td>
</tr>
<tr>
<td>CA</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>TX</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>&lt;1%</td>
<td>2%</td>
</tr>
<tr>
<td>CA</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>NC</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

*Not reported

Those of drivers associated with general crashes. These authors concluded that, despite the observed relationship between certain variables and fatal crash involvement, "it should not be inferred that we can accurately predict which specific individuals will be involved in a fatal accident" (6).

- Point systems that weight past violations and/or crashes have improved prediction of individual crash rates. In a study of point systems in the state of Washington, for example, it was reported that no particular type of violation was more likely than any other to predict subsequent crash involvement, with the exception of a few violations that were cited only rarely and hence had little predictive utility (5).

Based on these observations, the same researcher studied three alternative systems of weighting violations for the purpose of predicting crash involvement: Washington State’s severity point system, a simple count of total violations, and a point system based on collision probabilities associated with different violations as observed in the earlier research. Age, sex, and number of prior crashes also were included as predictors. The analyses indicated that none of the point systems predicted crash involvement better than did knowledge of the driver’s age and sex. Moreover, information on prior violations and crashes without age and sex information permitted no better than chance prediction of subsequent crash involvement.

In a more elaborate study, researchers in California compared the effectiveness of four schemes of predicting crash involvement over a three-year period (7). The first was California’s standard point system. The second emphasized five types of traffic violation patterns associated with nighttime crashes in California. The third and fourth schemes were based on regression analyses using either two variables (total convictions and total crashes) or four variables (total convictions, total crashes, age, and sex). The least effective scheme was found to be the standard California point count. The two regression models were better predictors. The other scheme fell between the point count and the regression models and was not significantly different in effectiveness from either.

Similarly, it was reported that neither of two point systems in California improved the ability to identify drivers who were subsequently involved in fatal crashes (6). It was not reported whether either of these systems or the total count of violations identified drivers involved in fatal crashes in the absence of information about age and sex.
Summary

No method currently exists to predict with accuracy the crash involvement of individuals, even when crash records are augmented by other variables such as age and sex. As mentioned earlier, one factor in this low level of predictability is the huge variation in police accident reporting procedures. However, even with better measures of crash experience, the predictive accuracy for individuals would remain low. Some researchers have suggested that even the best predictive variables theoretically could explain no more than 10 to 15 percent of the variation in three-year crash experience because of the large size of the random component in individual crash likelihood (10).

This is not to say that there are not identifiable groups of drivers with elevated crash risk. Problem drivers with very deviant records are several times more likely to crash than other drivers, and barring them from driving will eliminate some crashes. However, these people account for such a small segment of the problem that preventing them from driving can have little direct effect on overall crash totals. Thus, for actions such as license suspension or revocation to be effective in reducing highway crashes, they must serve as effective deterrents to other drivers. Whether they are effective as deterrents has yet to be determined.

References


Figure 1
Reported Crashes for Two 2-Year Periods — North Carolina
(Data from Stewart and Campbell, 1972)

(a) Percent of drivers with at least one crash in second two years, by number of crashes in first two years.

(b) Percent of crashes in second two years accounted for by drivers with specified number of crashes in first two years.

(c) Percent of drivers with specified number of crashes in first two years.