A Factor Analysis of Accident Records

TERRENCE M. ALLEN
Michigan State University

If statistical summaries of traffic accidents published yearly by most states and by the National Safety Council are interpreted in the obvious way, conclusions are likely to be erroneous because the age or sex of the driver may be attributed to extraneous factors such as the proportion of driving done in rural areas where high speeds are common, or to similar factors such as night driving, conditions that prejudice the results in terms of the drinking driver, and related characteristics of the driver, his vehicle, and his environment. For this reason, the factor analysis procedure has proved to be useful as a means for better understanding of a multivariable situation, as the traffic accident is.

In this study, eight major factors have been analyzed for their contribution to accidents. In addition, a separate factor analysis was conducted on a random sample of 1,000 fatal accidents. An elementary linkage analysis was also supplied to the data, primarily to group individuals into various classes, and this adds confidence to the results obtained.

It was found that most of the common variance among the 23 variables studied is accounted for by 8 independent factors. The report on this investigation explains the nature and the strength of the association of these eight factors with the accident event. It is quickly admitted that the inclusion of still further variables could result in other factors being identified with traffic accidents, as well as with a better definition of the ones that were isolated. It is important to recognize that particular combinations of a number of variables are probably of much greater importance in understanding the accident event than the effects of variables taken singly or in pairs.

•ALTHOUGH most people are aware that it is erroneous to think of accidents as having a single cause, most of the analysis of accident data is congruent with a single-cause theory of accidents. Consider the yearly statistical summaries of traffic accidents that are published annually by most states, and published for the nation by the National Safety Council. These tabulations convey the magnitude of specific problems, such as accidents involving excessive speed, alcohol, or pedestrians, and convey changes in the general accident problem and specific accident problems from year to year. However, it is difficult to make valid inferences regarding causation from such tabulations of one variable at a time. In addition, such summaries often include cross-tabulations conveying the relationship between two factors of interest, such as sex of drivers and number of accidents involving high speeds or alcohol. Although we may be aware that if we interpret such tables in the obvious way our conclusions are likely to be erroneous—sex of driver may be related to extraneous factors, such as the proportion of driving done in rural areas where high speeds are common, or to factors such as the proportion of driving done at night when drinking is more common—out methods of analysis typically do not take into account more than two variables at at time. Similarly, our studies relating characteristics of drivers, vehicles, and highways to accident rates typically treat variables two at a time. With such methods, only interpretations based on a

Paper sponsored by Committee on Highway Safety Research and presented at the 43rd Annual Meeting.

single-cause theory (tempered by judgment) are possible. There is need for the application of methods of multivariate analysis which are congruent with a multiple-causation theory of accidents. In the same way that the development of punched-card methods made it possible to record and tabulate data on many variables, the availability of modern computers makes possible multivariate analysis of such data.

Only a few applications of multivariate methods have been reported. Multiple-correlation and partial correlation methods have been used to a limited extent, such as to study the effects of roadway characteristics on the accident rates of sections of highway in Oregon (1, 2). Goldstein and Mosel (3) factor-analyzed attitude items and related the factors to self-reported accidents and violations, and Versace (4) factor-analyzed a portion of the Oregon data. There seems need for a multivariate attack on the large quantity of data available in accident records. The purpose of this study is to make a small beginning in that direction.

PROCEDURE

Accident Data

With the cooperation of the Michigan State Police, punched-card records were obtained on the fatal and injury accidents in Michigan for 1957, over 18,000 in total. Since it was impossible to analyze all of the data recorded for such accidents, 23 variables were chosen such that each accident could be coded by the presence or absence of each of 23 characteristics, and such that the characteristic not be so rare that very few accidents would be coded for it, or so common that almost all accidents would be coded for it. A conversion program was prepared so that the computer would make a card for each accident, recording only the data on these 23 variables. The characteristics are given in Table 1.

Reduction of the data to this form certainly leaves out much information of importance. In addition to data not recorded at all, the data on the variables chosen were not complete. For example, drivers' age was reduced to a simple dichotomy, whether driver under 25 was involved or not; information regarding ages over 65, or the age of the other driver, if any, was lost. A more complete analysis to recover such information is planned in a later study.

Factor Analysis as a Method

The principal multivariate method used in this study was a factor analysis of the correlation coefficients among the 23 variables which had been chosen and coded. This procedure was developed in the 1930's by Thurstone (5) and others for the study of psychological abilities. The basic idea of factor analysis can be illustrated by its most important early application. A large number of tests existed measuring various aspects of mental ability. Most of these tests had high correlations with some of the others. It was reasoned that there may be only a small number of basic factors of mental ability, and that each of the many tests was measuring one or more of these factors. Factor analysis was a means of deriving such factors inductively from the correlations among the tests. A small number of such factors were found, which made a basis for the further development of the theory of mental ability. The various intelligence tests in use today, and the interpretation and use of these tests, has been heavily influenced by factor analysis. Factor analysis has since been applied to diverse areas, and has proved its usefulness both in providing a means for better understanding of the variables involved, and in the practical development of efficient measuring instruments.

Correlation Matrix

The data with which factor analysis begins are given in Table 2. The table of the correlation coefficient of each variable with every other variable is known as a correlation matrix. In the case of dichotomous variables, the correlation coefficient reduces to what is called a phi coefficient, ϕ. It is related to the chi-square statistic,

TABLE 1

ACCIDENT CHARACTERISTICS STUDIED

1. Female driver involved or not
2. Driver under age 25 involved or not
3. Driver with less than one year's experience involved or not
4. Fatal accident or not
5. Alcohol involved or not
6. Accident at intersection or not
7. More than one vehicle involved vs single vehicle
8. Vehicle other than passenger car involved or not
9. Speed greater than 50 involved or not
10. Vehicle defect recorded or not
11. Out-of-state vehicle involved or not
12. Vision obscured or not
13. Daylight or not
14. Weather clear or cloudy vs other weather
15. Surface dry or not
16. Paved road or not
17. Road defect recorded or not
18. State or U. S. highway, or not
19. Traffic control devices present or not
20. Open country vs built-up or urban area
21. Weekend (5 pm Fri. 12 pm Sunday) or not
22. Rush hour (7-9 am or 4-6 pm) or not
23. Summer or not

χ^2, commonly used to test for a significant relationship between two dichotomous variables, as follows: $\phi = \pm \sqrt{\chi^2/N}$, where N is the sample size, and the sign of ϕ reflects a positive or negative relation between the variables.

To interpret Table 2, observe for example the top of column 1, which corresponds to "Female Involved." We see that accidents in which a female is involved are slightly less likely to involve a driver under 25; accidents involving a female are less likely to have a mention of alcohol on the accident record; they are more likely to be accidents involving more than one vehicle; they are more likely to involve an inexperienced driver, etc. The coefficients in Table 2 are based on 17,400 accidents (the number remaining after cards with incomplete data were eliminated, and when the number had been further randomly reduced to facilitate feeding cards into the computer in batches of 200), so that even very small coefficients are statistically significant. Examination of the matrix may reveal relationships of interest, but it is beyond human capacity to comprehend the matrix as a whole. Factor analysis may be viewed as a means of summarizing all of these coefficients in a way that the mind can grasp.

Methods of Analysis

The use of factor analysis is not an exact science, and opinions differ on a number of questions of procedure. The author's approach to the use of factor analysis is an empirical one; results which hold up over several methods of analysis are accepted, while those that do not are regarded as questionable. Accordingly, additional analyses were carried out, and their results used to guide the interpretation of the primary analysis.

For the primary analysis, factors were extracted by the principal components method (6), and rotated by the quartimax method (7) and the varimax method (8). Since the two methods yielded almost identical results, only those for the quartimax method are reported. Since these methods require orthogonal (independent) factors, the biquartamin method (9) was also applied. Again, almost the same solution was obtained. Since criteria for the number of factors to include in the rotation process did not lead

TABLE 2

INTERCORRELATION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1																							
2	-012																						
3	-148	-060																					
4	+111	+039	-083																				
5	+143	+068	-056	+470																			
6	+102	+207	-100	+024	+017																		
7	-073	+079	+107	-075	-045	-005																	
8	+168	-071	-304	+126	+131	+056	-039																
9	-026	+023	+050	+040	-024	+020	+119	+069															
10	+011	+018	-045	-036	-093	+047	-030	+037	+035														
11	-027	+027	+058	+031	-011	+019	+108	+043	+603	+029													
12	+006	-036	+045	+063	+123	-043	+046	-013	-022	-270	+016												
13	+009	-007	-070	+022	+025	+002	-052	+027	-224	+038	-158	-058											
14	+009	-038	+043	+153	+210	-043	+069	-027	-045	-189	-004	+380	-058										
15	-050	-002	+037	-206	-134	-026	+212	-053	-028	+072	-014	-055	+002	-038									
16	-017	+017	+001	-005	-012	+031	-011	+011	+036	+015	+033	+005	-019	-012	+002								
17	-052	-012	+000	-048	-037	-034	+042	-054	+003	-023	+021	+015	+016	+035	+077	-016							
18	-016	-020	-027	+014	+111	-010	+041	+021	-007	-020	+011	+034	-005	+070	+081	+001	+033						
19	-053	-015	-063	+099	+270	-017	-041	+087	+001	-016	+025	+017	+016	+039	-016	+029	+011	+107					
20	+001	+039	-021	+011	+014	+038	+045	+120	+105	+038	+199	-030	-012	-040	+012	+008	-014	+045	-008				
21	+076	-042	-143	+061	+065	+006	-034	+342	-006	+026	-026	-012	+025	-019	-041	-018	-034	-005	+030	-019			
22	-067	+040	+110	-018	+015	-007	+023	-090	-015	-018	-030	+008	-001	+027	+034	-001	+005	-006	-088	+012	-064		
23	+002	-048	+014	+020	+204	-070	+129	-031	-081	-138	-033	+232	-039	+397	+116	-017	+043	+162	+054	+016	-016	+015	

to a clear decision in this case, solutions were obtained for six through ten factors. Although factors seven and eight were of a doubtful status on the basis of the additional analyses below, it was felt that the eight-factor solution was most meaningful. Since the factor loadings on the first six factors were almost the same regardless of the number included in the rotation or the rotation method used, these loadings are considered reliable.

Results

The results of the factor analysis are given in Table 3. The principal results are the factors, indicated by roman numerals. The entries in the table, the factor loadings, measure the correlation between each variable and each factor. The person not acquainted with factor analysis will not be seriously in error if he thinks of a factor as a cluster of accident characteristics which hang together over many accidents, and the factor loadings as a measure of the degree to which a particular characteristic hangs together with the cluster. A negative loading then measures the degree to which the absence of a characteristic is associated with the cluster. Factor loadings smaller than 0.05, and decimal points, have been omitted to make the table easier to read. Each factor is defined by the variables on which it has the highest loadings, either positive or negative.

The first factor characterizes features of the road where an accident took place—whether it was paved, whether traffic control devices such as signs or signals were present, whether it was a state or U. S. highway, and whether a road defect was reported. The smaller loadings on "vision obscured" is reasonable, since vision may be obscured by roadway features; likewise the small loading on speed. This factor differentiates modern high-type highways from secondary roads and streets and will be referred to as "good roads."

Factor II clearly refers to the weather in which an accident took place, with loadings on variables 9 and 11 which describe the weather. The smaller loadings on 13 (vision obscured), 20 (summer), and 7 (speed) are consistent with this interpretation. This factor will be referred to as "weather."

TABLE 3

FACTOR ANALYSIS OF TOTAL ACCIDENTS[a]

Characteristic	Factor[b]							
	I	II	III	IV	V	VI	VII	VIII
1. Female	09	07	-45	-	-18	21	-	-17
2. Age under 25	-06	-	14	09	05	73	-	-13
3. Alcohol	07	-13	55	-07	-	-24	26	-05
4. Intersection	08	-06	-11	68	-26	06	10	-16
5. More than 1 vehicle	19	-	-09	80	-	06	-	-10
6. Experience 1 yr	-	-	-11	-	-	73	-	-
7. Speed over 50	18	-21	-	-17	49	13	11	-19
8. Daylight	-	-10	-77	13	-	-05	-	13
9. Clear or cloudy	-	-85	-	-	-	-	-	-05
10. Road defect	55	-	-08	-	16	-	08	-
11. Dry pavement	-	-84	-	-	-	-	-	-
12. Paved road	73	-	-	-	-08	-	-	07
13. Vision obscured	-22	43	-07	14	09	-	10	-
14. Traffic control device	72	-	-	20	07	-	-	-
15. Open country	-	06	-	-26	67	-	-	-
16. Vehicle defect	-	-08	09	-	-	18	-	88
17. Fatality	-	-	12	-06	23	-	-30	-22
18. Out-of-state vehicle	-	-	-	28	49	-06	-08	12
19. Trucks, etc.	-10	-	07	55	20	-10	-44	17
20. Summer	-11	-30	-20	10	23	-	40	10
21. Rush hour	-	-	-63	-	-	-15	-	06
22. Weekend	-	07	18	05	-	-	73	-
23. State or U. S. highway	58	09	-	16	42	-07	-05	-

[a] Decimal points and coefficients less than 0.05 omitted for ease of reading.

[b] Factors are I = roads, II = weather, III = night, IV = conflict, V = rural, VI = youth = inexperience, VII = weekend, and VIII = vehicle defect.

The interpretation of factor III is not so obvious. Accidents characterized by this factor are at night (or dusk or dawn), not during the rush hour, alcohol is likely to be involved, but not female drivers (who presumably do more of their driving during the day). The smaller loadings fit into the pattern—summer because of longer daylight hours, weekends because of more night traffic on weekends. In interpreting factors it is advisable to refer back to the correlation matrix and to compare the correlations between variables with loadings on the factor. When this is done it is seen that even the small loadings, such as those on age less than 25, and fatality, are consistent with the overall pattern. Although one might consider a designation such as "time of day" for this factor, it seems that the narrower designation of day vs night characterizes this pattern of characteristics more accurately. This factor will be referred to as "night."

Factor IV characterizes accidents which involve more than one vehicle, are at intersections and involve a vehicle other than a passenger car (in most cases, a truck). We find smaller loadings on variables 18 (out-of-state vehicle) perhaps partly because many out-of-state vehicles are trucks, 15 (not in open country), 14 (traffic control devices present), and 23 (state or U. S. highway). Reference to the correlation matrix verifies this pattern; we see, for example, that truck accidents are more likely to be at intersections, and particularly likely to involve more than one vehicle. (Of course, whether the cause is trucks per se or their traffic exposure is not answered by these correlations.) Clearly this factor implies something like traffic friction among vehicles, or interference in traffic movement, rather than just traffic congestion. This factor will be designated "traffic conflict."

Factor V yields a quite clear interpretation, characterizing accidents taking place in open country rather than urban or built-up areas, with high speeds and out-of-state vehicles more frequently involved. The smaller loadings, as well, point to a designation of this factor as "rural."

Factor VI has large loadings only on age under 25, and experience less than one year. The small negative loading on alcohol reflects the fact shown in Table 1, that alcohol is less likely to be associated with inexperienced drivers. The loading on female similarly reflects the greater likelihood of females being associated with experience less than one year. For want of a better term, this factor will be referred to as "youth-inexperience." It would seem that further data would be needed (or even analysis of male and female drivers separately) in order to make a clear interpretation of this factor.

Factor VII seems to characterize "weekend" accidents. The relation between weekend and summer, trucks, and even alcohol, are verified by the zero-order correlations in Table 1. However, that of variable 17, fatality, is not so verified although this loading is almost identical for the verimax solution, and holds up over different rotations. It may be that when other factors related to fatalities and weekends are adjusted fewer fatal accidents will be found on weekends than would be expected. Until verified by appropriate analysis, however, such relation appears doubtful. One might expect a loading on out-of-state vehicles as well, but this presumably does not occur because a large number of trucks are out-of-state vehicles, and few of these are on the road on Sundays. Perhaps Sundays need to be treated separately rather than treating the weekend as a whole. This factor should be interpreted with caution. It accounts for less variance than the previous ones, and was included in the rotation because the other factors emerged more clearly with it than without it.

Factor VIII, "vehicle defect," reflects the fact that vehicle defects do not seem to be strongly related to any of the other variables. Including it in the rotation allowed it to emerge by itself without its small relationships slightly disturbing the definition of the other factors. The amount of variance associated with this factor was quite small, and vehicle defects are reported in only a very small portion of accidents.

Additional Analyses

Since fatal accidents are of particular interest, a separate factor analysis was carried out on a random sample of 1,000 fatal accidents. (A sample of 1,000 non-fatal accidents was also analyzed. However, since about 94 percent of all accidents

TABLE 4

FACTOR ANALYSIS OF FATAL ACCIDENTS[a]

Characteristic	Factor[b]							
	I	II	III	IV	V	VI	VII	VIII
1. Female	-	-	-55	-	-	10	07	-
2. Age under 25	-	-09	23	12	-	61	-07	-41
3. Alcohol	05	-	24	-	08	-16	63	16
4. Intersection	-10	-14	-14	76	-18	-	10	-
5. More than 1 vehicle	21	05	-10	78	16	06	-08	-
6. Experience 1 yr	-08	-	-08	-	-	73	-	16
7. Speed + 50	19	-26	-05	-05	23	35	30	-06
8. Daylight	-11	-	-71	16	06	-	-18	06
9. Clear or cloudy	-	-85	-	-	-05	-	-	-
10. Road defect	-50	09	-10	08	20	09	07	-18
11. Dry pavement	-	-83	10	06	10	06	-	-
12. Paved road	70	-11	07	-08	-06	-	-08	-
13. Vision obscured	-16	51	11	15	08	23	-10	-
14. Traffic control device	75	-	-	11	-	-	10	-
15. Open country	06	-	-21	-25	64	15	-	-25
16. Vehicle defect	-	-	-09	-11	06	30	-	78
17. Fatality	-	-	-	-	-	-	-	-
18. Out-of-state	07	-06	05	27	57	-11	-12	10
19. Trucks, etc.	06	-	14	37	16	-	-54	32
20. Summer	-20	-20	15	12	48	-14	09	14
21. Rush hour	-	09	-66	15	-	-08	-	-
22. Weekend	-	08	09	13	-	17	62	-
23. State or U. S. highway	64	13	-	18	36	-09	-	-

[a]Decimal points and coefficients less than 0.05 omitted for ease of reading.
[b]Factors are I = Roads, II = weather, III = night, IV = conflict, V = rural, VI = youth = inexperience, VII = weekend, and VIII = vehicle defect.

were non-fatal, the results were essentially redundant with that of the total sample.) The results for the fatal accidents are reported in Table 4. It is obvious that the factorial structure is very similar to that of the total sample. The general description of the factors based on the total sample would apply almost equally well to the fatal sample. The pattern of loadings is very similar for factors I through VI, but not very close for factors VII and VIII. As the similarity of the factor analysis results implies, the correlation matrix for the fatal accidents was similar to that for the total sample given in Table 2; therefore, the matrix is not included for the fatals. Only a few of the correlations are sufficiently different to be of particular interest. The correlations of experience less than one year showed some suggestive changes from total sample to fatal—with speed, from -0.005 to 0.106; with alcohol, -0.100 to -0.027; with weekend, -0.007 to 0.072. Age less than 25 showed changes in the same direction, although to a lesser degree. The suggestion of a difference in the pattern of relationships between these variables for fatal vs non-fatal accidents points out the need for further analysis, taking account of the relations of these variables to others.

Since the mathematical model of factor analysis assumes linear relationships between continuous variables, its application to dichotomous variables introduces both theoretical and practical problems. Although it has proved useful in the analysis of such data (such as with test items which are answered right or wrong), the effect on the phi coefficient of differences in marginal proportions can lead to difficulties (10). One way that has been used to overcome this difficulty is to use ϕ/ϕ_{max} as a coefficient, where ϕ_{max} is the maximum ϕ possible for the given marginals. Samples of 1,000 fatal accidents and 1,000 non-fatal accidents were factored using this index. Again, the same factors were obtained, and in general the relative magnitudes of the loadings were similar. Since this index has undesirable mathematical properties that can lead to anomalous results in certain cases (11), the original analysis is preferred since the results are in general the same.

Another method of analysis, quite simple computationally and very different in mathematical model, was also applied. Elementary linkage analysis (12), although intended primarily to group individuals into types on the basis of agreement scores, can be used to group variables into types on the basis of their intercorrelations. This

method, applied to the matrix of the total sample, gave six types corresponding to factors I through VI. The variables in each type corresponded to the variables having the highest loading on the corresponding factor.

These secondary analyses, while adding relatively little to computer time, added considerably to the writer's confidence in the results. Other methods, more congruent with the dichotomous nature of the data and with other implicit hypotheses about its structure, are being tried out. To date, none has yielded a really satisfactory solution.

DISCUSSION

It seems clear that most of the common variance among the 23 variables can be accounted for by eight independent factors. Or, to put it informally, the 23 accident characteristics hang together in eight independent clusters.

Factor I, "good roads," is made up of characteristics in which modern high-type roads differ from secondary roads and streets. Although speed and multiple-vehicle accidents are related to this factor, it, to a great extent, stands by itself with a clear interpretation.

Factor II, "weather," although related to speed and time of year, is similarly clear in interpretation.

Factor III, "night," summarizes many characteristics on which the frequency of accidents is different at night than during the day. Most of the accident characteristics that are intrinsically of interest—age, sex, alcohol, speed, fatalities, etc.—are related to this factor. Although this factor may be viewed as a conglomerate—further research may characterize better the ways in which day and night accidents differ—it seems reasonable to think of this factor as a basic characteristic of accidents.

Factor IV, "traffic conflict," is characterized by intersections, more than one vehicle, and to a lesser extent, trucks. It is also related to variables of interest, and, more than the previous factor, might be viewed as a conglomerate that future research may clarify. But certainly this group of characteristics is important in the study of accidents. Some may prefer some other term like traffic interference or friction to conceptualize this factor.

Factor V, "rural," characterizes accidents that take place in rural areas rather than in urban or built-up areas. This seems clearly interpretable as a basic characteristic of accidents.

Factor VI, "youth-inexperience," should not be interpreted as a basic factor in the same sense as the previous ones. Except for sex, age under 25 and experience less than one year were the only purely human characteristics among the 23 included, so it is natural that they cluster together. This factor is of interest, however, and the suggestion of a different pattern for fatal and non-fatal accidents warrants further research.

Factor VII, "weekend," is of a very doubtful status. It was "underdetermined," i. e., it had too few variables with high loadings to determine it accurately, although its relation to alcohol at least is of importance. Lumping the whole weekend together may be erroneous—Sunday is certainly different for trucks, and probably for other characteristics.

Factor VIII, "vehicle defect," tells us little, except that vehicle defects are not strongly related to any other accident characteristic included in the study. Therefore, it came out pretty much by itself.

Interpretation of Factors

It is important to point out what these results may mean in practice, and what they do not mean. As in other applications of factor analysis, the results may help in conceptualizing the problem. The author, as a psychologist, is primarily interested in the human factors in accidents. However, the characteristics which might be regarded as human characteristics did not come out in factors by themselves, but clustered with non-human characteristics as well. It seems clear that study of human characteristics must take account of related non-human characteristics.

It was pointed out in the introduction to this paper that studies comparing accident rates or accident characteristics may lead to erroneous conclusions because of failure to control for extraneous variables. It is impossible to control for everything, and it can be quite confusing to decide what to control for and how control may be achieved. The factors obtained in this study are at least some important ones, and variables of interest may be related to some of the accident characteristics included in this study. By examination of the factor loadings (especially accompanied by examination of the correlation matrix), one may be able to identify some major variables and find some ways of control for them. For example, studies involving alcohol should control for at least the factors alcohol was found to be related to: night, youth, weekend, and perhaps weather. Similarly, if one were to compare summer accidents to those occurring the rest of the year, it is seen that this characteristic is related to seven out of the eight factors, making conclusions about summer accidents very likely to be influenced by extraneous variables.

Misinterpretation of Factors

It should be pointed out that these factors do not represent basic causes of accidents; they concern relationships among accident characteristics, and say nothing directly about causation. It cannot be concluded that these are the most important factors regarding accidents (as will be seen below, Factor I may be of minor importance); the factors represent merely clusters of characteristics commonly recorded in accident records. (If hair color, eye color, and skin color had been included in the analysis they would have clustered together as a factor, even though they are obviously not of importance.)

Also, these factors only summarize the data included in the analysis. The characteristics included in this study were judged to be the most important ones which could be obtained unambiguously from accident records in a form suitable for factor analysis. Including more characteristics would presumably result in additional factors in addition to better definition of the ones obtained.

Further Research Needed

It seems clear that this study should be regarded as a beginning. The need for the analysis of further variables, and further information on the variables studied, seem obvious. Also, factor analysis only takes account of the correlation between pairs of variables. It is reasonable to hypothesize that it is particular combinations of several variables that are important, over and above the effects of variables singly and in pairs. It is hoped that fruitful analysis methods, within the feasibility of computers, will be found to analyze accident data in ways consistent with this hypothesis and with other reasonable hypotheses.

Only one other study has been reported with which results can be compared. Versace (4) did a factor analysis pertinent to accidents. However, the units of analysis in his study were sections of highway and his variables were characteristics of the highway, one of which was an index of the number of accidents on each section. He found four factors which accounted for most of the common variance between highway characteristics: capacity, modern roads, traffic conflict and roadside structures. His factor, traffic conflict, which appears to correspond to the factor of the same name in this study, accounted for almost all the accidents variance. His factor, modern roads, which seems to correspond to the factor good roads in this study, was not related to accidents. Of course, it is only conjecture to equate the factors in the two studies. His modern roads factor was defined by high loadings on sight distance restriction (negative), calculated capacity, type of terrain, and number of curves (negative). His factor, traffic conflict, had its high loadings on accidents, average daily traffic, number of intersections, and number of commercial and residential driveways. Only one variable was common to the two studies--number of intersections per mile in one, and whether the accident took place at an intersection in the other study.

The need for variables common to the two types of study seems obvious. In studies of the accident characteristics of roadways, data on the proportion of truck traffic and

speeds, as well as certain accident characteristics such as were included in this study, would seem indicated. Similarly, it would be desirable to include on each accident card some code which would make it possible for a computer to make an estimate of the traffic volume, and the data necessary for calculated capacity, since these seem important with respect to where accidents occur. There is also need for related studies in which the unit of analysis is drivers, rather than accidents or sections of roadway. When we have a set of factors (or some other statistical model) which holds up over these three methods of analysis, we will have a basis on which a theory of accidents can be built.

ACKNOWLEDGMENTS

The research reported in this paper was initiated with support of the Highway Traffic Safety Center at Michigan State University. The research was carried on with support by National Institute of Health Grant AC-30. Gratitude is also expressed to the Michigan State Police who supplied the accident records, and to Charles Kiesler, Donald Wilkins, and James Clark for assistance with computer programming and data-processing.

REFERENCES

1. Schoppert, D. W. Predicting Traffic Accidents from Roadway Elements of Rural Two-Lane Highways with Gravel Shoulders. Highway Research Board Bull. 158, pp. 4-26, 1957.
2. Blensly, R. C., and Head, J. A. Statistical Determination of Effect of Paved Shoulder Width on Traffic Accident Frequency. Highway Research Board Bull. 240, pp. 1-23, 1960.
3. Goldstein, L. G., and Mosel, J. N. A Factor Study of Drivers' Attitudes, with Further Study on Driver Aggression. Highway Research Board Bull. 172, pp. 9-29, 1958.
4. Versace, John. Factor Analysis of Roadway and Accident Data. Highway Research Board Bull. 240, pp. 24-32, 1960.
5. Thurstone, L. L. Multiple-Factor Analysis. Chicago, Univ. of Chicago Press, 1947.
6. Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. J. Educ. Psychol. Vol. 24, pp. 417-441, 498-520, 1933.
7. Neuhaus, J. O., and Wrigley, C. F. The Quartimax Method: An Analytic Approach to Simple Structure. Brit. J. Stat. Psychol., Vol. 7, pp. 81-91, 1954.
8. Kaiser, H. F. The Varimax Criterion for Analytic Rotation in Factor Analysis. Psychometrika, Vol. 23, pp. 187-200, 1958.
9. Harman, H. H. Modern Factor Analysis. Chicago, Univ. of Chicago Press, 1960.
10. Wherry, R. J., and Gaylord, R. H. Factor Patterns of Test Items and Tests as a Function of the Correlation Coefficients: Content, Difficulty and Constant Error Factors. Psychometrika, Vol. 9, pp. 237-244, 1944.
11. Comrey, A. L., and Levonian, E. A Comparison of Three Point Coefficients in Factor Analysis of MMPI Items. Educ. Psychol. Meas., Vol. 18, pp. 739-755, 1958.
12. McQuitty, L. L. Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies. Educ. Psychol. Meas., Vol. 18, 1958.