

Factor Analysis of Pedestrian Accidents

L. V. Dickinson, Jr.,* MITRE Corporation
J. W. Hall, Transportation Studies Center, University of Maryland

Factor analysis was used to investigate more than 1000 pedestrian accidents reported in Maryland from 1970 to 1973. The Maryland accident record system was searched to select 120 sections of roadway that were potentially hazardous to the pedestrian. Data collected at these study sites from field examinations and photologs were then combined with data obtained from the accident record system and subjected to factor analysis. The factors that contribute to making certain locations hazardous to pedestrians relate to traffic conflicts, road conditions, nighttime illumination, the physical condition of both the pedestrian and the driver, and the pedestrian's lack of regard for his own safety. The usefulness of factor analysis in identifying hazardous locations is illustrated with a hypothetical example.

One of the most urgent problems facing the traffic engineer today is how to provide safe and efficient pedestrian movement. The interaction between pedestrians and vehicles on or adjacent to roadways results in delay and a less orderly flow for both types of traffic. Past operational practices seem to have assigned a higher priority to keeping the vehicles moving, giving concomitantly less emphasis to the pedestrian. The feeling was that, if the flow of vehicles could be maintained, the pedestrian would somehow fit into the environment and make his own way. However, with increased volumes of vehicular and pedestrian traffic, conflicts between pedestrians and vehicles have increased and forced those concerned with the community, including the engineer, planner, politician, and psychologist, to consider or reconsider techniques to promote a safer mixing of modes.

Accident data presented by the National Safety Council (1) indicate that pedestrian injuries and fatalities tended to decrease between 1937 and 1961. This trend was reversed and pedestrian accidents increased from 1962 to 1972. Despite a 2 percent decrease in 1973, the 10 500 pedestrian fatalities in that year accounted for almost 19 percent of all highway fatalities. In comparison with

nonpedestrian traffic fatalities, the percentage of pedestrian fatalities that occurred in urban areas was considerably higher (64 percent versus 25 percent). National statistics also indicate that more than 54 percent of the pedestrian fatalities occur at night. Of the 2.1 million traffic injuries reported each year, approximately 7 percent are received by pedestrians. In Maryland, of the more than 328 000 traffic accidents reported in a 4-year period (1970 to 1973), fewer than 2 percent (6268) involved pedestrians. These comparatively small percentages are misleading since injuries to pedestrians are generally more severe than those to occupants of vehicles.

A more realistic indication of the problem is given by the severity index, the ratio of fatal plus injury accidents to total accidents. In Maryland the severity index is 0.32 for all traffic accidents and 0.99 for pedestrian accidents. In other words, virtually all reported pedestrian accidents involve a fatality or an injury. Another indicator of the seriousness of pedestrian accidents is the fatality index, the ratio of fatal to total accidents. The fatality index for pedestrian accidents in Maryland is 0.083, approximately 10 times higher than the index for all other accidents. These characteristics are summarized in Table 1.

The objective of the research reported in this paper was to evaluate pedestrian accidents in order to develop effective countermeasures. The study procedure included the use of Maryland's accident record system to identify sections of roadway in Maryland that have conditions that were apparently hazardous to pedestrians. The statistical techniques of factor analysis were subsequently used to identify and investigate the factors that are associated with pedestrian accidents.

To aid in understanding the pedestrian problem and in determining the elements that contribute to pedestrian accidents, the data were dealt with in five categories: the driver, the environment, the pedestrian, the roadway, and the vehicle. A failure or breakdown within any of these five categories could contribute to a pedestrian accident. This pedestrian accident, described as a pedestrian-vehicle conflict, was defined as the actual accident resulting from a collision between one or more pedestrians and vehicles.

Publication of this paper sponsored by Committee on Pedestrians.

*Mr. Dickinson was with the University of Maryland when this research was performed.

SELECTION OF STUDY SITES

Before collecting our data, we had to select study sites that were potentially hazardous to the pedestrian. Using three criteria, we determined these sites from the accident data contained in the Maryland Department of Transportation accident record tapes for the period 1970 to 1973. First, we identified sections of roadway that had a high number of recorded pedestrian accidents per kilometer of roadway type within each county and determined the countywide ratio of pedestrian accidents to total length of roadway of each type. Second, we looked for pedestrian accidents that clustered along certain routes, since in some instances there were recurring accidents at the same location, although the number of accidents per kilometer was not significant. Third, we looked for consistent identifying characteristics of pedestrian accidents on certain types or sections of roadway, e.g., midblock accidents, bus-stop-related accidents, accidents involved with turning.

Computer programs were written to identify on the record tapes accidents that involved pedestrians and to create, for each of the 4 years, tapes that contained only pedestrian accidents. The three criteria were then applied to the records on the shorter tapes to select the study sites. This process yielded a total of 120 sections of roadway selected for field examination. Although the study sites were located in 16 of Maryland's 23 counties, most of the study sites were in the more urbanized counties, which had higher frequencies of pedestrian and vehicle conflicts. The study sites are identified below according to type of roadway.

Type of Roadway	Percentage of Maryland Roadways	Number of Study Sites
Interstate	1.4	3
U.S.	3.6	24
State	17.7	77
County	69.2	16
Municipal	8.1	0
Total	100.0	120

COLLECTION OF DATA

To make certain that the maximum amount of data was obtained at each of the study sites, a pedestrian field survey form was developed and filled out for each of the 120 sections of roadway. An example of this form, showing the type of information collected, is presented in Figure 1.

Before the study sites were investigated in the field, we located each accident site by using a listing that noted the milepost number for each accident. At the reported location of each pedestrian accident, we observed conditions that may have contributed to the conflicts. This information was used later in recommending remedial action or policy changes designed to improve the unsafe environment. The conditions we observed included improper use of shopping center entrances, driveway exiting problems, poor location of crosswalks and bus stops, restricted sight distance, extremely wide intersections, misuse of pedestrian crosswalks, poor or nonexistent sidewalks, and extremely poor maintenance. These conditions were recorded on the pedestrian field survey form and photographed for future reference.

After visiting about half of the accident sites, we concluded that filming the sections of roadway would be very useful since it would permit us to view and review the study sections on film, accident by accident. With the assistance of the Maryland Department of Transportation, we filmed approximately 483 km (300 miles) of

roadway on one-directional photologs to capture the 101 study sections that were located on state and U.S. highways. Although the photologs posed several problems, they did provide the opportunity for a detailed examination of each accident location. The film presented the situation that would normally be observed through the driver's eye. It also permitted us to obtain measurements directly from each film frame.

FACTOR ANALYSIS

The data collected in this study were treated under three headings: general pedestrian accident characteristics within Maryland, pedestrian accident characteristics at the study site locations, and factor analysis of the pedestrian accident data. The analysis and results of the first two are presented in a separate report (2); the third is discussed here.

The primary emphasis in previous analyses of accident data has involved tabulations and cross tabulations of two or more variables. Although this type of analysis is necessary, it promotes the theory of a single cause of accidents. Studies have shown that there are many interrelated variables that contribute to accidents; summary statistics, annual reports, and tabulations indicate the scope of pedestrian accidents rather than provide an adequate description. Thus, there is a need to apply the methods of multivariate analysis that are congruent with a multiple-causation theory of accidents. Factor analysis, one of these techniques of multivariate analysis (3, 4), was used in this study. This statistical procedure has been used extensively in social science research (5, 6). Like any other statistical procedure, it is a valuable tool if used correctly.

The most significant virtue of the technique with respect to this work is its ability to explain the relationship among many variables. For example, if a predictor variable correlates with a criterion variable, factor analysis tells "why" the correlation exists, in terms of which factors account for the correlation. Thus, certain relationships are explained in terms of more basic concepts, which in turn help to explain the overall problem.

In this research, factor analysis was used to search for meaningful relationships among the many variables associated with pedestrian accident data. By determining the relationship between groups of variables and more basic factors, we can achieve a clearer understanding of the pedestrian problem. The greater insight into the problem provided by this technique can then be used to develop realistic types of countermeasures or remedial actions for reducing hazardous conditions and establishing a safer environment for the pedestrian.

Data on the pedestrian accidents at the study sites, together with information on the design and operating characteristics of the roadways, were analyzed by using the BMDO3M, a factor-analysis computer program (7). A detailed analysis was performed on the 243 pedestrian accidents that occurred during the 4-year period at the 26 study sites in Prince George's County, near Washington, D.C. Using data from the accident record system, we categorized 1090 pedestrian accidents at the 120 study sites according to five types of land use. The six classifications used and the number of variables analyzed for each are identified below.

Classification	Variables Analyzed
Prince George's County	23
Manufacturing-industrial land use	21
Shopping-business land use	22
Residential land use	22
School-recreational land use	20
Open country	22

Table 1. Selected Maryland accident statistics (excluding Baltimore City).

Year	All Traffic Accidents	Pedestrian Accidents							
		Total							
		Num-ber	Per-cent	Fatal	Fatal-ities	In-jury	In-juries	Severity Index ^a	Fatality Index ^b
1970	74 005	1366	1.8	114	116	1249	1288	1.00	0.083
1971	82 650	1584	1.9	136	138	1437	1512	0.99	0.086
1972	79 950	1603	2.0	139	139	1435	1513	0.98	0.087
1973	91 030	1715	1.9	132	132	1538	1588	0.97	0.077

^a(Fatal + injury accidents)/total pedestrian accidents.

^bFatal accidents/total pedestrian accidents.

Table 2. Variables and their values.

Number	Description	Value	Number	Description	Value
1	Lane width	Numerical	13	Good weather (versus bad)	Dichotomous
2	Number of lanes	Numerical	14	Daytime (versus night)	Dichotomous
3	Sidewalk width	Numerical	15	Weekend (versus weekday)	Dichotomous
4	Median width	Numerical	16	Summer (versus nonsummer)	Dichotomous
5	Speed limit	Numerical	17	Dry surface (versus not dry)	Dichotomous
6	Commercial driveways per kilometer	Numerical	18	Good road (versus bad)	Dichotomous
7	Residential driveways per kilometer	Numerical	19	Pedestrian normal (versus not normal)	Dichotomous
8	Intersections per kilometer	Numerical	20	Driver normal (versus not normal)	Dichotomous
9	Average daily traffic of vehicles	Numerical	21	Intersection (versus nonintersection)	Dichotomous
10	Parking (versus no parking)	Dichotomous	22	Peak period (versus off-peak)	Dichotomous
11	Average daily traffic of pedestrians	Numerical	23	Accident score ^c	Numerical
12	Shoulder width	Numerical			

^cAccident score: 3 = fatal, 2 = injury, 1 = property damage only.

Figure 1. Pedestrian field survey form.

Section No. 96 Date 11-6-74

County Prince George Rt. No. 70 214 Rt. Name Central Ave.

Length of Section 0.85 Milepost 1.17 to 2.02 ADT 17,000

Roadway Surface Bituminous Roadway Width 20' # Lanes 2

Lane Width 10' Shoulders Yes Shoulder Width 8'

Sidewalk No Sidewalk Width - Curbs No Lighted No

Median No Median Width - Delineation Centerline

Vertical Curvature * Horizontal Curvature None

Landuse Residential/Commercial Speed Limit 40 Slide Nos. B-1 (15, 17, 20)

Intersections per Section 7 # Intersections per Mile 8

Total # Driveways/Section 35 Total # Driveways/Mile 41

Commercial Driveways/Section 10 # Commercial Driveways/Mile 12/30

Parking: One Side - Both Sides - None ✓

Accidents Within Section 10 # Recurring Accidents 1(5), 1(6)

Milepost of Recurring Accidents 1.32 (1.33) 1.34 ; 1.36 (1.37) 1.38

Milepost at Start of Section Begins at Addison Road @ 0.87

Comments:

* Crest vertical curve at recurring accident sites.
Very poor pavement markings

Factor analysis was applied separately to each of the categories.

The variables used in the investigation were obtained from both the coded accident records and the field studies. Approximately half of the variables had numerical values, while the remainder were dichotomous. The use of dichotomous values was dictated by the nature of some of the variables, which had to be coded as 0 for a favorable condition or 1 for an unfavorable condition. Previous research (5, 8) has used this dichotomous structure to increase the sensitivity of factor-analysis programs in detecting variations within the data. The vari-

ables that were analyzed are shown in Table 2.

The initial factor analysis was conducted by using data for the study sites in Prince George's County. This was the only area in which information was available for variable 11, pedestrian average daily traffic (ADT). This variable was important because, as volumes of pedestrian and vehicular traffic increase, the potential for accidents also increases. With assistance from county personnel, the pedestrian ADT was obtained for each of the 26 study sites within Prince George's County.

The correlation matrix for the 23 variables, which is used as input to the factor-analysis program, is shown in Table 3. The coefficients appear reasonable in both sign and magnitude. For example, variable 2, number of lanes, is positively correlated with variable 9, vehicle ADT, and negatively correlated with variable 7, residential driveways per kilometer. These correlations suggest that the accidents occurred on roadways that have good design features. These facilities are characterized by higher vehicular speeds, greater volumes of traffic, and lower numbers of residential driveways. Variable 11, pedestrian ADT, is positively correlated with variable 3, sidewalk width, and negatively correlated with variable 5, speed limit. Not surprisingly, variable 13, good weather (versus bad), is highly correlated with variable 17, dry surface (versus not dry). Variable 19, pedestrian normal (versus not normal), is not significantly correlated with any of the remaining variables.

Factor Matrix

The correlation coefficients shown in Table 3 were used to develop a factor matrix. This matrix shows the correlation between each of the 23 variables and the developed factors, which were selected by the computer program in a stepwise manner, depending on the amount of variance between variables that is explained. For example, the first factor selected would explain the maximum variance between each of the variables, while the second factor, orthogonal to the first, would explain the largest possible amount of the remaining variance. Each factor has an eigenvalue, which is the sum of the variance of each variable explained by a factor. The process continues until all the variance is explained. At this point the number of independent factors equals the original number of variables.

Interpretation of a matrix this size is difficult because of the number of elements. To simplify the interpretation, eigenvectors that explain an insignificant amount of variation can be eliminated. In addition, it is often worthwhile to rotate the remaining factors to ensure a more meaningful interpretation. Since a certain amount of explained variation for each variable is lost by elimination of some of the factors, rules have been established for determining which factors should be selected for rotation (3, 9, 10). The rotated factor matrix for the pedestrian accident data is presented in Table 4, which shows the coefficients of correlation between each variable and factor. These coefficients are also referred to as factor loadings. Although it is not shown in the table, the total variance of each variable (communality) remains constant during rotation of the factor matrix.

Describing the Factors

The traffic engineering interpretation of the rotated factor matrix can best be illustrated through an example. Factor I has the highest eigenvalue. The loadings on this factor varied from +0.88 to -0.75. The larger positive loadings indicate that certain variables were highly positively correlated with factor I; i.e., the accident locations were characterized by wider shoulders, medians, and lanes and by higher speed limits. Several variables exhibit a strong negative association with the accident data, indicating that the locations were characterized by infrequent occurrence of intersections and commercial driveways, limited availability of sidewalks, and low volumes of pedestrian traffic. Table 4 also indicates that there was little correlation between factor I and the remaining variables.

The variables that were negatively correlated with factor I provide as much information about the accidents as those that were positively correlated. The finding that certain variables (11, pedestrian ADT; 6, commercial driveways per kilometer; 3, sidewalk width; and 8, intersections per kilometer) were negatively correlated with this factor suggests that these accidents occurred away from urban or built-up areas. This is supported by the finding that other variables (12, shoulder width; 5, speed limit; and, to a lesser degree, 4, median width; and 9, vehicle ADT) were positively correlated with factor I. In view of these conditions, the factor was described as open roadway without pedestrian facilities (a description not unique to us), primarily on the basis of the factor loadings. The loading of the accident score also indicates the existence of more severe injuries among these accidents, which would occur in the more open suburban areas with lower traffic volumes and a lesser number of points of interference between the pedestrian and vehicle.

A similar approach was followed in characterizing and identifying the nine remaining factors. Although the second factor also refers to features of the roadway, this factor is characterized by multilanes, wider medians, higher speed limits, and higher vehicle ADT. It is also characterized by a negative loading on variable 7, residential driveways per kilometer. Factor II was thus described as pedestrian-restricted high-speed roadway.

The third factor clearly refers to the weather at the time the accident occurred (note the extremely high loadings found on variables 13 and 17). Factor III was referred to as weather.

The fourth factor is characterized by higher loadings on variables 14, daytime (versus night), and 22, peak periods (versus off peak), which is consistent with the correlation found between those two variables (Table 3). Factor IV was therefore described as traffic density.

For this factor, the accident score had a loading of 0.30, reflecting a greater severity of injury during peak periods.

Factor V had unusual factor loadings. Its negative association with medians (variable 4), as well as a high correlation with parking (variable 10) and somewhat smaller loading for weekend accidents (variable 15), seemed to relate this factor to a restricted or impeded area, such as a vehicle terminal area. The lack of peak-period accidents and the positive loading for pedestrian ADT strengthen this association. Therefore, factor V was described as roadway impedance (friction).

The sixth factor was associated with a high positive loading on lane width (variable 1) and smaller positive loadings on both pedestrian ADT (variable 11) and the condition of the pedestrian (variable 19), which indicate association with some type of crossing. The negative loading on variable 8 (intersections per kilometer) suggests that these crossings occurred away from intersections. With this in mind, factor VI was described as pedestrian crossing.

The seventh factor clearly represents the alcohol problem. This interpretation is a result of the negative loading of variable 19 and the high positive correlation of variable 20. Factor VII was described as the drinking pedestrian.

The eighth factor presented a problem in interpretation. It was associated with nighttime accidents that occurred during nonsummer months (variables 14 and 16) and showed a tendency for the presence of residential driveways and the absence of commercial driveways (variables 7 and 6). Analysis of the correlation matrix (Table 3) substantiated the description of factor VIII as season.

Factor IX was associated with a high loading on variable 18, good road (versus bad), apparently indicating that these accidents largely occurred on good pavement surfaces. The factor loading for variable 15 also showed a higher occurrence of weekend than weekday accidents. Neither the factor matrix nor the correlation matrix assisted in establishing the exact nature of this factor. The most reasonable description of factor IX was road condition.

Factor X had the lowest eigenvalue above the selected base of 1.0 and was the most difficult to describe. The three most notable characteristics were the positive loading for commercial driveways per kilometer (variable 16), the high positive loading for intersection versus nonintersection (variable 21), and the negative loading for pedestrian ADT (variable 11). The first two reflect the merging, diverging, and crossing of traffic streams. Assigning major importance to these two variables, we described factor X as potential conflict points.

Approximately 75 percent of the variance among the original 23 variables was explained by these 10 independent factors.

Factor	Description
I	Open roadway without pedestrian facilities
II	Pedestrian-restricted high-speed roadway
III	Weather
IV	Traffic density
V	Roadway impedance (friction)
VI	Pedestrian crossing
VII	Drinking pedestrian
VIII	Season
IX	Road condition
X	Potential conflict points

Although some of them were easily described, several factors, especially those with low eigenvalues, were difficult to describe because they contained few variables

Table 3. Correlation coefficients for variables.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	1.00																						
2	0.33	1.00																					
3	-0.32	0.08	1.00																				
4	0.24	0.28	-0.26	1.00																			
5	0.16	0.49	-0.47	0.24	1.00																		
6	-0.07	0.30	0.19	-0.19	-0.11	1.00																	
7	0.01	-0.60	0.02	-0.27	-0.47	-0.45	1.00																
8	-0.36	-0.04	0.49	-0.29	-0.55	0.44	0.04	1.00															
9	0.19	0.70	-0.10	0.53	0.68	-0.02	-0.54	-0.26	1.00														
10	0.30	0.32	-0.22	-0.29	0.38	-0.02	-0.18	-0.06	0.18	1.00													
11	0.05	0.16	0.32	-0.27	-0.39	0.20	-0.01	0.19	-0.19	0.14	1.00												
12	0.19	0.23	-0.56	0.39	0.52	-0.50	0.05	-0.60	0.34	0.04	-0.40	1.00											
13	0.15	0.05	-0.11	-0.06	0.11	-0.05	-0.01	-0.21	0.02	0.15	0.04	0.06	1.00										
14	-0.02	0.04	0.04	-0.01	0.09	0.05	-0.11	-0.01	0.08	-0.03	—	—	0.10	1.00									
15	0.15	0.10	-0.05	-0.13	0.11	0.03	-0.07	-0.01	0.12	0.23	0.05	0.04	0.06	-0.03	1.00								
16	-0.01	0.02	0.01	-0.04	0.03	0.09	-0.11	-0.02	0.01	-0.06	0.06	-0.05	0.01	0.18	-0.04	1.00							
17	0.17	0.07	-0.15	-0.04	0.13	-0.05	0.01	-0.23	0.05	0.13	—	0.08	0.87	0.09	0.08	0.01	1.00						
18	0.04	0.02	0.05	-0.05	-0.03	0.04	0.04	-0.02	0.05	0.07	0.01	0.02	0.10	-0.02	0.08	-0.05	0.12	1.00					
19	0.10	0.02	-0.02	0.08	0.02	0.09	-0.12	-0.04	-0.02	-0.07	0.07	0.01	-0.05	0.20	-0.03	0.09	-0.01	-0.09	1.00				
20	—	—	0.11	-0.03	-0.08	-0.07	0.05	0.12	-0.02	-0.01	0.10	-0.08	-0.01	0.16	0.06	0.03	-0.02	-0.03	-0.08	1.00			
21	0.03	0.06	-0.15	0.07	0.17	0.05	-0.03	-0.12	0.13	-0.03	-0.27	0.14	-0.12	0.07	-0.03	-0.03	-0.04	0.03	-0.01	—	1.00		
22	-0.04	0.02	0.07	0.03	—	0.03	-0.04	0.02	0.03	-0.06	0.04	-0.05	0.03	0.47	-0.10	-0.05	0.02	-0.08	0.17	0.01	0.04	1.00	
23	0.07	0.02	0.09	-0.09	0.11	-0.13	-0.06	-0.13	0.14	0.01	0.01	0.13	—	0.10	0.07	-0.01	-0.02	0.05	0.08	0.06	-0.01	0.10	1.00

Note: A dash in the matrix denotes zero correlation.

Table 4. Rotated factor matrix.

Variable	Factor									
	I	II	III	IV	V	VI	VII	VIII	IX	X
1	0.23	0.12	0.11	-0.05	0.13	0.85	0.04	0.04	0.02	0.05
2	-0.20	0.86	0.06	-0.02	0.12	0.31	0.06	0.07	-0.02	0.01
3	-0.07	—	-0.06	0.05	-0.18	-0.20	0.19	0.08	0.10	-0.20
4	0.38	0.46	-0.09	-0.04	-0.63	0.24	0.06	0.08	0.04	-0.01
5	0.61	0.63	0.07	0.07	0.29	-0.12	-0.07	-0.04	-0.06	0.13
6	-0.65	0.30	—	0.02	0.09	—	-0.32	-0.22	—	0.26
7	0.11	-0.82	0.02	-0.08	-0.05	0.09	0.19	0.20	0.02	-0.01
8	-0.75	-0.11	-0.19	—	0.02	-0.24	0.05	0.03	-0.02	0.01
9	0.29	0.86	-0.01	0.02	-0.04	0.02	0.11	0.08	0.11	0.02
10	0.09	0.20	0.10	-0.03	0.82	0.18	—	0.14	-0.11	-0.02
11	-0.52	-0.06	0.03	0.04	0.18	0.37	0.07	-0.04	—	-0.44
12	0.88	0.03	—	-0.01	-0.06	-0.02	-0.01	0.02	0.08	0.05
13	0.08	0.01	0.94	0.05	0.06	0.03	—	-0.01	0.03	-0.10
14	-0.01	0.06	0.11	0.79	0.01	-0.05	0.17	-0.22	0.02	0.11
15	0.04	0.08	-0.02	-0.08	0.53	0.07	0.10	-0.06	0.34	-0.05
16	-0.01	-0.04	0.03	—	-0.03	-0.02	0.07	-0.90	-0.05	-0.03
17	0.10	0.02	0.94	0.05	0.04	0.07	-0.02	-0.02	0.06	-0.01
18	-0.09	-0.01	0.18	-0.14	0.03	0.01	-0.05	0.08	0.80	0.19
19	—	-0.01	-0.13	0.43	-0.11	0.36	-0.44	-0.28	0.02	-0.10
20	-0.10	-0.03	-0.04	0.13	0.03	0.07	0.85	-0.12	—	0.02
21	0.12	0.05	-0.12	0.11	-0.02	0.06	0.05	0.03	0.09	0.83
22	-0.06	0.02	0.04	0.82	-0.07	-0.02	-0.03	0.23	-0.09	0.01
23	0.24	0.09	-0.17	0.30	0.06	—	0.06	-0.02	0.51	-0.37

Note: A dash in the matrix denotes a zero factor loading.

Table 5. Specific values and factors selected for hypothetical location.

Variable	Value	Factor Selected	Variable	Value	Factor Selected
1	3.7 m	I, VI	13	0	III
2	6	II, VI	14	1	VIII
3	—	V, VI, X	15	0	V, IX
4	6.1 m	I, II, VI	16	1	VIII
5	88.5 km/h	I, II, V	17	0	III
6	—	I, VII	18	0	IX
7	—	II	19	0	IV, VI
8	1.9	I, III, VI	20	0	VII
9	30 000	I, II	21	0	X
10	1	IX	22	1	I, V, IX
11	25	I, X	23	3	I, IV, IX
12	3.1 m	I			

Note: 1 m = 3.3 ft; 1 km = 0.6 mile; variables 10 and 13-22 are dichotomous.

with high loadings. The absence of more definitive information made their description less meaningful.

The factor-analysis techniques were subsequently applied to pedestrian accident data for all 120 study sites classified by land use. Information on pedestrian ADT was not available for study sites outside of Prince George's County, so this variable was not included in the analysis. The findings of this analysis are presented in a separate report (2).

Using the data from the 26 study sites in Prince George's County, we found that wider shoulders, higher

speed limits, wider medians, greater numbers of lanes, and higher vehicle ADT characterized accident locations in nonurban areas. The absence of sidewalks, residential driveways, commercial driveways, intersections, and pedestrians substantiated these interpretations.

To apply these findings to subsequent accident investigations, one must consider both the factors and their associated variables. For example, two related areas of concern, based on recorded accident data, were factor I, open roadway without pedestrian facilities, and factor II, pedestrian-restricted high-speed roadway. The analysis showed that both factors were negatively correlated with pedestrian ADT. For a pedestrian accident to occur, however, a pedestrian must have been present. Since specific accidents lose their identity in the factor-analysis process, it was not possible to determine whether the accidents associated with factors I and II occurred on controlled-access roadways. If they did, stricter enforcement is required to keep pedestrians off these roadways. If they occurred on other roadways, better pedestrian safety education programs, coupled with engineering countermeasures, may be warranted.

As we expected, variables that have direct application to engineering solutions clustered around factors I and II. The lack of engineering-related problems may also suggest that other variables should be included in future investigations. If the factor analysis reflected a clustering

of variables concerned with horizontal or vertical curvature around a particular factor, this might indicate that locations with poor sight distance contribute to the occurrence of pedestrian accidents. Further investigation could be made at such locations and, if necessary, countermeasures developed and implemented that could be as minor as removing excessive foliage or shrubbery, restricting pedestrian movement, or providing advance warning to motorists or as major as a complete redesign and construction.

Factors characterized by variables related to the environment or human behavior demonstrate the need for other types of action to provide for a safer pedestrian environment. Cases in which either the pedestrian or the driver was not in normal condition and that also exhibit positive correlation with such other variables as daytime (versus night) and summer (versus nonsummer), as in factor VIII, indicate a need for nonengineering forms of remedial action. These might involve more stringent enforcement of the law concerning both the drinking driver and the drinking pedestrian, as well as safety education programs.

Factor analysis identifies the variables that are associated through many incidents. An examination of these variables leads to the interpretation of a particular problem area. By investigating these problem areas, it is possible to determine which variables in the field should be considered in developing adequate solutions. Taking this process one step further, factor analysis can be used for improving specific locations. For example, assume that recorded accident data indicate that a particular location in Prince George's County is potentially hazardous. The information presented in this study could then be used to determine what the problems are and what effort can best be made to solve them. By using the 10 factors developed from the Prince George's County factor analysis, it may be possible to categorize this particular location under one or more of these factors. Engineering personnel could then select the variables that were highly correlated with this interpreted factor, and these variables would warrant special consideration in planning improvements for the specific area.

Applying the Factor Analysis

A simplified scheme for applying the findings of this research would initially consider the values of the 23 original variables. The average values (both numerical and dichotomous) of these variables would be compared with those at the specific location under consideration. Table 5 presents values for a hypothetical location. The lane width at this location is 3.7 m (12 ft). Since the average lane width for the accident data was 3.4 m (11.2 ft), a factor from Table 4 would be selected that had a high loading on variable 1. Factors I and VI fit this criterion for selection. The number of lanes at the specific location (6) is also larger than the average for the accident data (4.6). Again, we select a factor that has a high loading on variable 2. This criterion is met by factors II and VI. The third line in Table 5 shows that there is no sidewalk at our hypothetical location. Factors V, VI, and X have high negative loading on variable 3, indicating the negative association between pedestrian accidents and sidewalks. This same process is continued for the remaining 20 variables (average values for Prince George's County data for the numerical variables: 4 = 3.5 m, 5 = 56.3 km/h, 6 = 17.5, 7 = 11.1, 8 = 8.1, 9 = 26 730, 11 = 460, 12 = 0.6 m, 23 = 2.05), including the dichotomous variables. For example, since the value for variable 22 shows that the accidents occurred during the off-peak periods, we select factors I, V, and

IX, which have negative loadings for this condition.

Table 5 identifies the factors selected for each of the 23 variables. The selection frequency for each factor is then determined. In this example, factor I was selected 10 out of the 23 times and factor VI was selected 6 times. The remaining factors were each chosen less frequently. Since all of the variables have equal weighting, the specific location would then be classified under factor I, open roadway without pedestrian facilities. Table 4 should be reexamined, considering the variables that were highly loaded on factor I, to determine their relevance to the specific location. The appropriate variables can then be used to develop countermeasures for reducing hazards to pedestrians.

CONCLUSIONS

This study determined that there were many factors that make certain locations hazardous for pedestrians. Most of these unsafe conditions were related to areas of traffic conflicts, nighttime illumination, and intoxicated pedestrians and drivers. Even though implementation of the recommended countermeasures would help create a safer pedestrian environment, a stricter observance of existing controls is an integral part of this safer system. A characteristic common to many of the accidents was the pedestrian's lack of attention to his own safety, as reflected by the fact that nearly 80 percent of the reported accidents cited the pedestrian as the cause. The pedestrian's attitudes must be altered so that the maximum benefit can be obtained from engineering. The forms of remedial action that the engineer can implement on a large scale can only ameliorate, rather than solve, the problem of pedestrian accidents.

The factor analysis used in this research was shown to be a useful method for studying the problem of pedestrian accidents. The solutions developed in the original study (2) were based on the results of factor analysis along with a knowledge of traffic conditions at the locations. Because factor analysis lies somewhere between a science and an art, different researchers may interpret the results differently. Therefore, it is not appropriate to develop solutions to the problem of pedestrian accidents by using statistical techniques alone. Traffic engineering continues to play an important element in the planning and implementation of remedial action for the specific conditions surrounding hazardous locations. The use of factor analysis simply directs the engineer's attention to the combinations of roadway conditions that are most closely associated with pedestrian accidents.

To increase the effectiveness of factor analysis, especially in this area, which is closely involved with individual behavior, it is recommended that future research include variables that are concerned with the human factors. Analyzing this type of data might assist in explaining the factions of pedestrians who are involved in accidents. An examination of the human element would be beneficial not only in research but also in the interpretation of accidents that involve a drinking pedestrian or driver.

Further research should also include a comparable number of variables from each of the five elements involved in pedestrian accidents—the driver, the environment, the pedestrian, the roadway, and the vehicle. In this study, 14 of the 23 variables that were investigated pertained to the roadway. Future analysis should consider reducing the number of roadway variables and increasing those for the four remaining areas. The analyst should thus be in a better position to develop broad-based but workable methods of promoting pedestrian safety.

The potential for using the techniques of factor analysis to develop hazard indexes warrants further attention.

This research investigated only those locations with unusually high occurrences of pedestrian accidents. Future research should also examine locations that were not found to be hazardous. In this way the characteristics of both safe and unsafe locations could be analyzed and compared, so that indexes could be established to determine the relative hazard to pedestrians at different locations. These indexes could then be used to formulate a priority listing of unsafe locations so that corrective measures can be taken to alleviate hazardous pedestrian conditions.

REFERENCES

1. Accident Facts. National Safety Council, Chicago, 1974.
2. L. V. Dickinson, Jr. Factor Analytical Approach for Identifying and Solving the Pedestrian Problem. Univ. of Maryland, College Park, PhD dissertation, May 1975.
3. H. H. Harman. Modern Factor Analysis. Univ. of Chicago Press, 1960.
4. L. L. Thurstone. Multiple Factor Analysis. Univ. of Chicago Press, 1947.
5. T. M. Allen. A Factor Analysis of Accident Records. HRB, Highway Research Record 79, 1965, pp. 17-25.
6. J. Versace. Factor Analysis of Roadway and Accident Data. Chrysler Corp., Detroit, 1960.
7. BMD03M Factor Analysis. Biomedical Computer Program, Univ. of California, Berkeley, 1966.
8. M. C. Nizlek. Pedestrian Traffic Accident Characteristics: Tucson, Arizona. Arizona Transportation and Traffic Institute, Tucson, Research Project 74-613-002, June 1974.
9. W. W. Cooley and P. R. Lohnes. Multivariate Data Analysis. Wiley, New York, 1971.
10. D. F. Morrison. Multivariate Statistical Methods. McGraw-Hill, New York, 1967.