Transportation and Traffic Engineering, Univ. of California, Los Angeles, Rept. Rr-68-63, 1968.
9. R. Henson, C. Cannell, and S. Lawson. Effects of Interviewer's Style and Question Form on the Validity of Reporting Automobile Accident Information. Highway Safety Research Institute, Univ. of Michigan, Ann Arbor, HIT Lab Reports, Vol. 3, No. 12, Aug. 1973.
10. P.J. Cooper. Effects of Increased Enforcement at Urban Intersections on Driver Behavior and Safety. TRB, Transportation Research Record 540, 1975, pp. 13-22.
11. J. Byun, W.R. McShane, and E.J. Cantilli. Changing Baseline in Transportation Safety: An Assessment of Some Key Factors. TRB, Transportation Research Record 709, 1979, pp. 6-10.
12. Alabama Municipal Data Book, 1980. Alabama Office of State Planning and Federal Programs, Montgomery, 1981.
13. Information Bulletin, Director of Mayors and Commissioners in Alabama. Alabama League of Municipalities, Montgomery, Aug. 1980.
14. The Alabama Municipal Journal. Alabama League of Municipalities, Montgomery, Vol. 38, No. 8, Feb. 1981, pp. 14-18.
15. Crime in Alabama: 1979. Uniform Crime Reports Division, Alabama Criminal Justice Information Center, Montgomery, 1980.
16. Urban and Rural Accident Statistics. Accident Identification and Surveillance Section, Alabama Highway Department, Montgomery, 1975-1979 Computer Summaries.
17. Alabama County Data Book, 1980. Alabama Office of State Planning and Federal Programs, Montgomery, 1981.
18. R.F. Gunst and R.L. Mason. Regression Analysis and Its Applications. Marcel-Dekker, Inc., New York, 1980.
19. O.F. Dunn and V.A. Clark. Applied Statistics: Analysis of Variance and Regression. Wiley, New York, 1974.
20. S. Chatterjee and B. Price. Regression Analysis by Example. Wiley, New York, 1977.
21. E.R. Mansfield, R.F. Gunst, and J.W. Wester. An Analytic Variable Selection Technique for Principal Component Regression. Journal of the Royal Statistical Society, Series C, Vol. 26, 1977, pp. 34-40.

Publication of this paper sponsored by Committee on Traffic Records.

Application of Microcomputer Technology to Local Accident Problem Identification

DAVID B. BROWN AND CECIL W. COLSON

The potential for implementing a microcomputer-based problem identification system in small to medium-sized cities is explored in terms of the City Accidents RAPID Evaluation (CARE) system. The benefits of such a system are examined. One primary benefit is overall data improvement for all applications. The capabilities of CARE are explained in terms of a user-oriented menu-driven operating system. Example outputs are presented along with the methodology for their generation. Finally, some technical specifications are provided to illustrate considerations required for actual installation.

Problem identification is an essential part of the design of an optimal safety system at all levels (1). NHTSA has recognized the criticality of performing systematic problem identification and has incorporated this as a requirement for each state highway safety plan (HSP) (2,3). But problem identification is also essential at the local level for local countermeasure implementation. In fact, the closer to the source of the problem the process of problem identification and evaluation is performed, the more effective it will be. For example, if local law-enforcement officers knew the locations in their city where accident rates are high as well as the times and types of accidents at those locations, they would then be in a position to implement selective enforcement countermeasures.

The benefits of having a local problem identification capability are obvious. Being able to obtain information for specific problem subject areas (such as accidents involving alcohol or pedestrians) or specific locations gives the local decisionmaker the information needed to develop an optimal allocation of resources. A few years ago, it was not economically feasible to provide direct on-line query capability to a small town. Now, however, with the advances made in microtechnology, every city and town of any reasonable size can take advantage of the tools that have been developed. One of the major benefits of distributed problem identification will be the tremendous increase in accuracy of the records themselves as local police realize the important role that accident records can play in countermeasure development.

City Accidents RAPID Evaluation (CARE) (4) is a microcomputer-based system that enables city officials to quickly retrieve information stored in their accident records. The users of CARE need no formal training in computer hardware or software since no knowledge of computers is required. The various options of CARE are incorporated into menus that thoroughly guide the user in obtaining the desired output. By following the directions given on these menus, all output required can be readily obtained at the terminal and/or on a printer.

CARE is patterned after Records Analysis for Problem Identification and Definition (RAPID) (5), a system developed for statewide accident problem identification that has been installed in Alabama, Kentucky, South Carolina, Tennessee, and Delaware. The differences between CARE and RAPID are as follows:

1. CARE is designed to operate on its own dedicated hardware, a microcomputer in the $10 000-$15 000 price range, whereas RAPID requires a large system because of the large subsets necessary for statewide application.

2. CARE is designed to provide problem identification for a moderate-sized city, whereas RAPID is generally applied statewide.

3. CARE will generally be based on the most recent three years of accident data, whereas RAPID problem identification generally uses one year of data. (Informal studies in Alabama have shown that three years provides an optimal balance between data completeness and timeliness.)

4. Because of the local application, variables used for statewide research will generally be excluded from CARE. The CARE variables are only those that have application to local decisionmaking.

CARE provides the following capabilities directly to the user:

1. The user can create a subset of the city's accident records according to any logical specification, such as all pedestrian accidents, alcohol-related fatal accidents, motorcycle accidents at a given intersection, etc.

2. The user can obtain labeled frequency distributions for the accident subset chosen. Note that the production of total citywide frequency distributions (for all variables) falls within this capability. Any or all variables (such as time of day, day of the week, weather, etc.) can be selected by the user on-line.

3. Labeled histograms (graphs) of any frequency distribution produced by capability 2 can be obtained.

4. Fully labeled cross tabulations for any of the variables for subsets produced by capability 1 can be obtained.

5. The user can find high accident locations according to user-specified criteria [a "location" is specified by road code(s) and/or mileposts]. The interactive nature of this task enables the user to try any number of alternative criteria in order to obtain the number and the type of high accident locations desired.

6. The user can obtain frequency distributions for any variable for locations found by capability 5 to be high accident locations or for any other location specified by the user.

7. The user can obtain the accident report numbers for any subset of the accident records so that the hard copy for particular types or locations of accidents can be retrieved.

Although the capabilities of CARE are quite sophisticated, they can be obtained by anyone by merely selecting an option on a menu. To understand the capabilities of CARE, it helps to follow the flow of data from the origin to the final CARE output report. Suppose that a pedestrian accident occurs in a specific city, injuring a child on his or her way to school. This accident is recorded on a standardized form by an officer of the local police department. This hard-copy form is sent to a central point in the state for data entry. It, along with thousands of other records, becomes part of the state's accident records data base, which is generally stored on tape.

The accident records data base described above is generally not constructed with problem identification in mind. In fact, it contains virtually all of the codeable elements from the accident records. Many of these are not required for problem identification work, and generally they are not in a form that is compatible with problem identification. For example, the pedestrian's actual age will probably be coded on the tape whereas age intervals (e.g., 0-4, 5-7, 8-9, 10-15, 16-21, etc.) would be much more useful for problem identification and cross tabulation. In addition, certain calculations and other data manipulation might be required for problem identification. For this reason, it is essential to reformat and modify this data base before trying to use it for problem identification.

To resume the trace of the data, a program is now run to create a new tape which is totally compatible with CARE formats and objectives. (The original tape remains unaffected and may be used for other purposes.) It is important to note here that CARE can work on any properly formatted data base. The arrangement, number, or type of variables is totally flexible and may be specified by the user during the development of the BASE program. At this point, the records for the particular city to be considered are selected from the reformatted tape and downloaded onto the microprocessor hard disk. This new subset of data that resides on the hard disk is called the CARE Master Data Base. For cities in the 20 000 to 40 000 population grouping, this will produce from 3000 to 7500 records for a three-year time period. For larger cities, fewer years of data will be sufficient for problem identification purposes.

Residing in a small subset on the microprocessor hard disk, the data are now ready for quick processing through any of the CARE processing options: frequencies, histograms, or cross tabulations. They can also be processed through high accident location determination and reporting modules.

For example, suppose that the pedestrian accident above has caused considerable pressure in a given city for an increase in the number of school crossing guards. In order that these additional guards be placed where and when they can be most effective, a report consisting of the age distribution of the pedestrian accidents in that city over the past three years by severity, day of the week, pedestrian sex, and pedestrian action is required. Further, the specific locations at which the 7- to 9-year-old age group had accidents and the times at which these accidents occurred will be useful in making tactical decisions. With CARE this information is readily available, and it can be printed out for a report in a matter of a few seconds after the request is made. Similarly, any other classification of accidents can be studied in detail by using CARE.

CARE OPERATIONS

An overview of the CARE functions can easily be obtained because CARE is totally menu driven. This means that option lists appear on the terminal that thoroughly guide the user in obtaining every CARE output. No supplementary documentation is required to use the system, and no knowledge of computers or data processing is required. To illustrate this concept, consider the CARE supervisory menu (SM) shown in Figure 1, which will appear on the terminal screen when the system is turned on. The top row of keys on the terminal are function keys labeled F1, F2, ... F11. The user merely pushes the key that corresponds to the desired function on the menu. The following summary illustrates the CARE capabilities.

SM-F1 - CARE Logic Specification

The SM-F1 option enables the user to specify any subset of records. Standard subsets such as pedestrian, speed, or alcohol will appear immediately as on the menu shown in Figure 2. The user can subdivide the data into any combination of these standard subsets--for example, speed and fatal accidents involving alcohol. The menu also gives the user the ability to combine subsets. For example, all accidents that are either bicycle or pedestrian related

Figure 1. CARE supervisory menu.

```
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
*                                                          *
*        F1 - CARE LOGIC SPECIFICATION                     *
*        F2 - FREQUENCY DISTRIBUTIONS, HISTOGRAMS          *
*        F3 - CROSSTABULATIONS                             *
*        F4 - FIND HIGH ACCIDENT LOCATIONS                 *
*        F5 - SPECIFIC LOCATION INFORMATION                *
*        F6 - DETERMINE ACCIDENT NUMBERS (ON)              *
*        F7 - BATCH FREQUENCIES (F2)                       *
*        F8 - BATCH CROSSTABULATIONS (F3)                  *
*        F9 - BATCH LOCATION INFORMATION (F5)              *
*        F10 - CARE MAINTENANCE MENU                       *
*        F11 - LOGOFF/RETURN TO CP/M                       *
*                                                          *
*   FUNCTION =>                                            *
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

Figure 2. CARE logic menu.

```
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
*                                                           *
*        __ ALL - ALL ACCIDENTS                             *
*           ALC - ALCOHOL RELATED ACCIDENTS
*           BIC - BICYCLE RELATED ACCIDENTS                 *
*           DRI - DRIVER RELATED ACCIDENTS
*           EMS - INJURY OR FATAL ACCIDENTS                 *
*           FAT - FATAL ACCIDENTS
*           MOT - MOTORCYCLE RELATED ACCIDENTS              *
*           PED - PEDESTRIAN RELATED ACCIDENTS
*           RDS - ROADWAY RELATED ACCIDENTS                 *
*           RRT - RAILROAD RELATED ACCIDENTS
*           SCB - SCHOOL BUS RELATED ACCIDENTS              *
*           TRK - TRUCK RELATED ACCIDENTS
*           VEH - VEHICLE RELATED ACCIDENTS                 *
*           YTH - YOUTH RELATED ACCIDENTS
*           SPC - SPECIAL USER DEFINED LOGIC                *
*                                                           *
*    F1-DOWN  F2-UP  F3-SELECT  F4-UNSELECT  F11-END        *
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

can be examined concurrently. All of this is performed by entering options specified at the bottom of the menu. Once the given subset of the records is determined, the user is queried for a title. This title will appear at the top of every page of output generated on the printer. After these specifications, the system returns control to the supervisory menu. The logic and the title specified will remain in effect until F1 is entered once again from the supervisory menu.

SM-F2 - Frequency Distributions, Histograms

Once the user specifies the subsets to be processed, the next step is to designate the processing desired. This option enables tabular summaries of any of the variables in the CARE master data base to be returned to the screen or printer in the form of frequency distributions. There is also the option to generate a graphical picture of any of these frequency distributions. If a hard copy of the variable list (the choice of variables and their numbers) is not available, the user can obtain a quick listing of these on the screen or the printer merely by pressing the appropriate function button given on the menu.

SM-F3 - Cross Tabulations

The SM-F3 option is another processing option that can be applied to any subset specified by SM-F1. When the cross-tabulation option is selected, the user will be queried for a pair of variables. The system will then print out the cross tabulation on the line printer, after which another pair of variables may be entered.

SM-F4 - Find High Accident Locations

The SM-F4 function is used to find high accident locations that may be hazardous or might need attention because of the volume of traffic. The objective of SM-F4 is to find locations that either have a high total frequency of all accidents or have a high frequency of a given type of accident (e.g., pedestrian, injury, or any other of the logical qualifications specified in SM-F1). The user is queried for the minimum number of accidents needed to qualify a location for this category. An entry of 1 will generate all locations. Once specific locations are found, frequency distributions and/or accident numbers for any of the locations can be generated by following the instructions given.

SM-F5 - Specific Location Information

The SM-F5 function is used to obtain detailed information on a location-by-location basis. It complements SM-F4, which finds the high accident locations. Once specific locations are found, the next logical step is to determine the details, such as weather, roadway condition, lighting, and time of day, at that location. When SM-F5 is selected, the user will be queried for the location to be processed. A selection may then be made of any of the variables in the data base. Abbreviated frequency distributions (i.e., without percentages) will be listed for these variables. This printout can then be used for an on-site investigation. Accident numbers can also be obtained within this option.

SM-F6 - Determine Accident Numbers

Accident numbers are required to obtain the hard-copy original of the accident report. These are of obvious value in investigating high accident locations. However, quite often specific information that has not been entered into the computer system will be required from the hard copy for a subset of accidents. For example, a recent request was made for the names of all children between the ages of 6 and 10 who were injured in pedestrian accidents in a given local area. CARE can be used to obtain the subset through logic specification (SM-F1). Once this has been specified, SM-F6 will return the accident report numbers for all accidents in the subset. Accident numbers can also be obtained by location within SM-F4.

SM-F7 - Batch Frequencies (F2)

The SM-F7 function enables frequency distributions and histograms to be generated on a production basis rather than on-line. The system will query for as many variables as the user desires to input. Variable ranges may be input as well as individual variables to save time in data entry. Once all variables are entered, the system will produce all frequency distributions and/or histograms requested.

SM-F8 - Batch Cross Tabulations (F3)

Cross tabulations can also be obtained in a batch mode to save user time. In this case, the user will be queried for all variables for which cross tabulations are required.

SM-F9 - Batch Location Information (F5)

In cases where several variables for a large number of locations are to be output, this batch option for specific location information will be used.

SM-F10 - Care Maintenance Menu

A number of "housekeeping" chores must be performed to keep the CARE system operating properly. For example, adding a new road code requires an addition to the road name file. Similarly, each year (or more frequently) new data will be added to the system and this requires the execution of certain utility programs. SM-F10 presents the user with an entire menu of these maintenance items.

SM-F11 - Logoff/Return to CP/M

The SM-F11 function makes possible a systematic shutdown of the system, which is necessary to protect data on the hard disk. It also makes possible a convenient exit to CP/M so that the hardware can be used for systems other than CARE.

EXAMPLE CARE OUTPUTS

Frequency Distributions, Histograms

When function F2 is entered from the supervisory menu, the CARE frequency menu will appear immediately. To specify a variable, the user merely inputs the variable number for which a frequency distribution is desired. The frequency distribution will appear on the screen in a matter of seconds. The listing on the screen is similar to that shown in Figure 3. (The example shown is for the city of East Greenford for the 1979-1981 period. The title input during SM-F1 was "Alcohol Injury Accidents Study," and the logic is for ALC and EMS--i.e., alcohol-related accidents that resulted in one or more injuries.) A message at the bottom of the screen tells the user to press any key to generate the histogram. When this is done, a histogram similar to that shown in Figure 4 will appear on the

Figure 3. CARE frequency analysis.

ALCOHOL INJURY ACCIDENTS STUDY 02-02-82
EAST GREENFORD: 1979-81
LOGIC SELECTION: ALC and EMS

C002 MONTH

CATEGORY LABEL	CODE	ABSOLUTE FREQUENCY	RELATIVE FREQUENCY (PERCENT)	ADJUSTED FREQUENCY (PERCENT)	CUMULATIVE ADJ FREQ (PERCENT)
January	1	2	4.4	4.4	4.4
March	3	4	8.9	8.9	13.3
April	4	6	13.3	13.3	26.7
May	5	2	4.4	4.4	31.1
June	6	9	20.0	20.0	51.1
July	7	2	4.4	4.4	55.6
August	8	1	2.2	2.2	57.8
October	10	3	6.7	6.7	64.4
November	11	11	24.4	24.4	88.9
December	12	5	11.1	11.1	100.0
	TOTAL	45	100.0	100.0	

Subset size: 45 Records (16.014% of 281 Record CARE Database)

Figure 4. CARE frequency histograms.

ALCOHOL INJURY ACCIDENTS STUDY 02-08-82
EAST GREENFORD: 1979-81
LOGIC SELECTION: ALC and EMS

C002 MONTH

```
CODE
    I
  1 ****** ( 2 )
    I  January
    I
  3 *********** ( 4 )
    I  March
    I
  4 **************** ( 6 )
    I  April
    I
  5 ****** ( 2 )
    I  May
    I
  6 *********************** ( 9 )
    I  June
    I
  7 ****** ( 2 )
    I  July
    I
  8 **** ( 1 )
    I  August
    I
 10 ******** ( 3 )
    I  October
    I
 11 ****************************** ( 11 )
    I  November
    I
 12 ************* ( 5 )
    I  December
    I---------I---------I---------I---------I---------I
    0         4         8        12        16        20
    Frequency
```

Subset Size: 45 Records (16.014% of 281 Record CARE Database).

Figure 5. CARE cross-tabulation analysis.

1981 SEVERITY ANALYSIS 02-25-82
EAST GREENFORD
LOGIC SELECTION: ALL

C006 ACCIDENT SEVERITY BY C040 LIGHT

C040	COUNT / ROW PCT / COL PCT / TOT PCT	PROPERTY DAMAGE 1	NONFATAL INJURY 2	FATALITY 3	ROW TOTAL
DAYLIGHT	1	435 84.6 65.9 54.1	77 15.0 55.0 9.6	2 0.4 50.0 0.2	514 63.9
DAWN	2	2 100.0 0.3 0.2	0 0.0 0.0 0.0	0 0.0 0.0 0.0	2 0.2
DUSK	3	19 82.6 2.9 2.4	4 17.4 2.9 0.5	0 0.0 0.0 0.0	23 2.9
DARKNESS	4	125 79.1 18.9 15.5	32 20.3 22.9 4.0	1 0.6 25.0 0.1	158 19.7
DARKNESS, RD. LIGHTS	5	79 73.8 12.0 9.8	27 25.2 19.3 3.4	1 0.9 25.0 0.1	107 13.3
	COLUMN TOTAL	660 82.1	140 17.4	4 0.5	804 100.0

SUBSET SIZE: 804 RECORDS (100.000% OF 804 RECORD CARE DATABASE).

Figure 6. CARE high accident locations.

BICYCLE ACCIDENT LOCATIONS
EAST GREENFORD 1979-1981 MINIMUM ACCIDENTS: 1
LOGIC SELECTION: BIC

ACCIDENTS	LOCATION-INTERSECTION ON	AT
2	0030-DEAN RD.	0035-AVONDALE ST.
1	0036-CLANTON AVE.	0047-BRUTON ST.
1	0043-COLSON LA.	0051-CLARK AVE.
1	0043-COLSON LA.	0067-CHURCH ST.

ACCIDENTS	LOCATION-NONINTERSECTION ON	BETWEEN	AND
1	0036-CLANTON AVE.	0047-BURTON ST.	0058-COOK ST.
2	0044-COX ST.	0018-BINFORD RD.	0019-BROWN ST.
1	0072-DELWOOD DR.	0021-BOWDEN AVE.	0038-CARTER ST.

TOTAL ACCIDENTS: 9

screen. The user is given the option to print none, one, or both of these outputs to the printer. Regardless of which of these options is chosen, the system then returns to the variable query and will follow through this same procedure with the next variable specified.

Cross Tabulations

When the cross-tabulation function is selected, the system immediately queries the user for the pair of variables for which cross tabulations are required. It then calculates the cross tabulation for the subset as defined in the logic specified in SM-F1. An example of cross tabulations is shown in Figure 5. Every attempt has been made to keep the CARE outputs to report-sized documents. The system will flip-flop variables and use compressed print when necessary to accomplish this end.

Find High Accident Locations

The SM-F4 function provides the mechanism by which high accident locations can be found. Whatever logic was specified within SM-F1 will form the basis for determining the type of location under consideration. After specifying whether intersection or nonintersection accidents are of concern, the user is immediately queried for the number of accidents that will determine a high accident location. This is followed by a query for printed or screen output. The system will then respond by listing all locations that meet all of the criteria specified.

Figure 6 shows the output generated from a typical high accident location run.

By following the directions given below the listing, the user can put the system into a "frequency" mode. In this mode, frequency distributions can be obtained for any of the locations on the screen. These are obtained by moving the cursor to the line listing the desired location and pressing the proper function key. The user will then be queried for the variable desired. Thus, any item of information for a specific location is at the fingertips of the user.

TECHNICAL SPECIFICATIONS

CARE can be applied in any political subdivision that has from a few hundred to tens of thousands of accidents per year. Hardware should be chosen to provide the degree of service desired. This will depend heavily on the number of records to be stored and processed. A minimal configuration consists of a 5-megabyte (5 million bytes) hard disk and a 64K central processing unit with terminal and printer. This will support a city that has approximately 2500 accidents/year, assuming three years of accident data are to be maintained on-line. The cost of such hardware is about $10 000. For a larger number of accidents and faster service, microcomputers of the 16-bit word length variety, which range in cost up to $30 000, can be configured.

Implementation of CARE for a political subdivision will require more than just the purchase of hardware and the installation of the CARE software. In addition, an interface program is required to establish the data base for the political subdivision under consideration. This is no great problem for those states that perform data entry of all recordable accidents. First, a program would be written to pick off and convert the appropriate records and variables from the state master tapes to the format required by CARE. This would be performed on a mainframe to minimize cost and save time. The second step would involve a downloading of the data from the mainframe to the micro. Standard programs for this are available and quite reliable.

In the case where the state data entry is not sufficient for the needs of the city, a program would be developed to perform the data entry locally. The expense of such a program is nominal compared with the cost of the ongoing data entry effort. However, many cities may opt for continual local data entry to ensure the integrity of their data.

CONCLUSION

Space does not permit all of the CARE features to be explained here. Rather, the objective is to demonstrate the reality of microcomputer-based systems for local problem identification. At the inception of the CARE system design and development project, the objective was to develop software on state-of-the-art hardware, looking some four or five years into the future when hardware development would produce a machine that was both fast enough and in the price range to make local problem identification feasible. When the software was developed and tested, it was found that there was no need for localities to wait: Hardware and software technology is now more than adequate for the task. This is not to say that there is no room for further innovation. Every attempt will be made to keep CARE current with the most recent advances in microtechnology.

ACKNOWLEDGMENT

Research for the system design and software described in this paper was funded by the Alabama Office of Highway and Traffic Safety.

REFERENCES

1. D.B. Brown. Systems Analysis and Design for Safety. Prentice-Hall, Englewood Cliffs, NJ, 1976.
2. D.B. Brown. Highway Safety Problem Identification. In Implementing Highway Safety Improvements, ASCE, New York, 1980.
3. Problem Identification Course: Data Analysis Requirements. NHTSA, April 1977.
4. RAPID: System Overview. Alabama Office of Highway and Traffic Safety, Montgomery, Rept. 300-78-002-401-095, 1978.
5. CARE: System Overview and User Guide. Alabama Office of Highway and Traffic Safety. Montgomery, Rept. 300-82-002-401-11, 1982.

Publication of this paper sponsored by Committee on Traffic Records.