

UC Berkeley

Research Reports

Title

High Collision Concentration Location: Table C Evaluation and Recommendations

Permalink

<https://escholarship.org/uc/item/9q17p6r6>

Authors

Ragland, David R.
Chan, Ching-Yao

Publication Date

2007-05-18

HIGH COLLISION CONCENTRATION LOCATION

TABLE C EVALUATION AND RECOMMENDATIONS

PREPARED FOR



Task Order 5215-6215

PREPARED BY

David R. Ragland
Traffic Safety Center (TSC)

Ching-Yao Chan
Partners for Advanced Transit and Highways (PATH)



Traffic Safety Center
Setting New Directions in Traffic Safety





Traffic Safety Center

Setting New Directions in Traffic Safety

University of California Traffic Safety Center ■ Institute of Transportation Studies

University of California ■ Berkeley, California 94730-7360

Tel: 510/642-0655 ■ Fax: 510/643-9922



DRAFT FINAL REPORT ■ MAY 18, 2007

HIGH COLLISION CONCENTRATION LOCATION

TABLE C EVALUATION AND RECOMMENDATIONS

PREPARED FOR



Task Order 5215-6215

PREPARED BY

David R. Ragland
Traffic Safety Center (TSC)

Ching-Yao Chan
Partners for Advanced Transit and Highways (PATH)

EXECUTIVE SUMMARY

This report describes the research work that was conducted under PATH Task Order 5215 and its extension Task Order 6215, “Methods for Identifying High-Concentration Collision Locations (HCCL).” The subject matter is related to regularly published Caltrans reports, so-called Table C, that are used to screen for and investigate locations within the California State Highway System that have collision frequencies significantly greater than the base or expected numbers when compared to other locations. The accuracy and reliability of such reports are critical as Table C is the basis for follow-up field investigation as well as potential safety improvements.

In recent years, a Caltrans Table C Task Force reviewed the practices of Table C and subsequently made recommendations for improvements based on the feedback from the users of such reports. Some immediate revisions were made to correct certain issue addressed by the Task Force, yet it was clear from the review that a more thorough research effort should be made to establish solid and sound methodologies to carry out the screening and identification of HCCL so that the overall execution of safety investigation and safety improvements performed over the California State Highway Network can be more efficient and consistent. This project arises from the needs to address these problems.

During the course of the project, the research team from the Traffic Safety Center and California PATH Program of University of California at Berkeley conducted extensive literature reviews and surveys, and interacted with a number of out-of-state agencies and experts to gather the latest information and techniques in dealing with the HCCL subject matters. The issues involving HCCL are broad as they cover a wide range of spatial and temporal parameters. Furthermore, the methodologies deserve to be investigated in depth as there are many mathematical and statistical details that may affect the outcome of the Table C process.

PRIMARY FINDINGS AND RECOMMENDATIONS

The primary findings and conclusions are summarized in the report with recommendations, when appropriate, for potentially addressing and improving the process of identifying HCCL. The findings and recommendations are organized by the respective nature or attributes of the issues into the following seven categories:

1 PHYSICAL STRUCTURE OF ANALYSIS UNITS—WHAT IS A SITE?

- Should analyses be conducted independently within Rate Groups, or should all categories of sites be compared together?
- If analyses are to be conducted within Rate Groups, how should Rate Groups be defined?
- For analyses of roadway segments, how should such segments be subdivided in the analysis?

2 TEMPORAL STRUCTURE OF ANALYSIS

- Length of time used to calculate the base rate
- Length of time used to estimate the risk at a specific site
- Frequency with which the analysis is conducted (e.g., quarterly, biannually, yearly?).

3 CHOICE OF OUTCOME(S)

- Weighting by level of severity (PDO, injury, fatality)
- Analyses by different collision types

4 CRITERIA FOR SELECTION OF LOCATIONS

- Table C method
- Safety Performance Function (SPF)

- Empirical Bayes (EB)
- Continuous Risk Profile (CRP)

5 FORMAT AND CONTENT FOR REPORTING SITES

- Information provided (e.g., highway factors, non-highway factors, collisions factors)
- Integrated Data System

6 DATA QUALITY

- Highway infrastructure
- Traffic volume
- Collision data

7 APPROACHES OTHER THAN SITE-SPECIFIC APPROACHES

- Individual Sites vs. Types of Sites
- Corridors

The recommendations made in this report for different categories of issues requires various levels of resources to execute or implement. Some minor problems can be tackled with no changes to the current Table C method with minimum programming efforts, such as fixing data errors or eliminating double counting of crashes. More involved issues require more in-depth evaluation and significant resources to implement, such as re-categorization of rate groups or adjustments of statistical approaches in screening HCCL. More specific observations and recommendations are given in the following pages.

1. PHYSICAL STRUCTURE OF ANALYSIS UNITS—WHAT IS A SITE?

1.1. SHOULD ANALYSES BE CONDUCTED INDEPENDENTLY WITHIN RATE GROUPS, OR SHOULD ALL CATEGORIES OF SITES BE COMPARED TOGETHER?

TABLE C:

Hierarchical organization with sub-categories of Intersections, Ramps, and Highways

Observations:

- Virtually all methods of identifying HCCLs conduct analyses within roadway categories.
- The result is “local” optimization but probably not “global” optimization.
- Causal factors and countermeasures vary substantially among different roadway categories.

Recommendations:

- Maintain general approach of conducting analyses within roadway categories.
- However, determine the impact of this approach on overall safety benefits.
- Develop a formal rationale and methodology for defining rate groups.

1.2. IF ANALYSES ARE TO BE CONDUCTED WITHIN CATEGORIES (I.E., RATE GROUP), HOW SHOULD RATE GROUPS BE DEFINED?

TABLE C:

Present structure is hierarchical with sub-categories of Intersections (30 subgroups), Ramps (80 subgroups), and Highways (67 subgroups).

Observations:

- The rationale for this particular structure has never been defined.
- Some of the rate groups have a very small number of member sites, leading to instability in determining HCCLs.

Recommendations:

- Establish a formal rationale for the structure of Rate Groups (several possible rationale are offered in the text).
- Establish a method for determining Rate Group structure (e.g., determine whether a similar statistical model can be used to define clusters of Rate Groups [e.g., whether rural, suburban, and urban four-way signalize intersections can be combined using the same statistical model]).

1.3. FOR ANALYSES OF ROADWAY SEGMENTS, HOW SHOULD SUCH SEGMENTS BE SUBDIVIDED IN THE ANALYSIS?

TABLE C:

A "sliding window" approach is used in which a 0.2 mile window is moved in increments of 0.02 miles.

Observations:

- There is a tradeoff in setting the window length: Peaks less than the width of the window will be masked; However, narrowing the window will create instability in both expected and actual collisions.

Recommendations:

- Develop more stable estimates of expected collision through use of Safety Performance Functions (SPFs) and Empirical Bayes methods.
- Develop methods for determining continuous risk profiles.

2. TEMPORAL STRUCTURE OF ANALYSIS

2.1. LENGTH OF TIME USED TO CALCULATE THE BASE

TABLE C:

Based rates are calculated using three years of data.

Observations:

- If a Rate Group has a sufficient number of member sites, three years ought to be sufficient to provide stable estimates of Base Rates; the variation in rates becomes fairly small with as few as 30-40 member sites in a Rate Group.

Recommendations:

- Maintain a three year period.
- Evaluate trends over an extended period of time to determine if there is "drift" in underlying Base Rates.

2.2. LENGTH OF TIME USED TO ESTIMATE THE RISK AT A SPECIFIC SITE

TABLE C:

Tests are conducted using 3 months, 6 months, or 12 months.

Observations:

- Any time period less than a year is too short to a stable estimate of HCCL, no matter how calculated. One year is adequate for sites with high volume but may not be adequate for sites with low volume.
- The current method used (and most other methods) are aimed at determining elevations in fixed risk at particular sites (i.e., assume that risk is constant over time). This method (and most other methods) is not designed to detect changes in risk over time.

Recommendations:

- Eliminate estimates based on any time period less than one year.
- Use a method proposed by Ezra Hauer to determine stability of estimates. Use that method to decide whether, for particular sites, the time period should be one year, or longer.
- Utilize a method recommended in SafetyAnalyst for determining changes in risk. This should be routinely applied to all sites on a quarterly basis, and especially in sites that are experiencing other changes.

2.3. FREQUENCY WITH WHICH THE ANALYSIS IS CONDUCTED (E.G., QUARTERLY, BIANNUALLY, YEARLY?)

TABLE C:

The report is issued quarterly.

Observations:

- The survey conducted by the Table C Task Force indicated that Caltrans users of the Table C report are in favor of having a quarterly report (as opposed to biannually or yearly).
- However, there is probably no purpose in calculating estimates of fixed risk more frequently than yearly.

Recommendations:

- The Table C report should be produced quarterly.
- However, a standard analyses for HCCLs should be done only on a yearly basis.
- Other quarters should include reports on sub-topics, particularly analyses of potential change in risk (see above).

3. CHOICE OF OUTCOME(S)

3.1. WEIGHTING BY LEVEL OF SEVERITY (PDO, INJURY, FATALITY)

TABLE C:

Treats collisions of all severities with equal weight

Observations:

- Many approaches to identifying HCCLs weight collisions by severity, with weighting increasing for PDO, injury, and fatality collisions.

- This approach has two major flaws:
 - If fatality is weighted too heavily it creates instability in the estimates, since fatality is rare
 - It assumes that collisions of different severity are similarly distributed across locations. In fact, PDO, injury, and fatal collisions have substantially different distributions. We have noted that the distribution of fatal and severe injury appear more closely related to one another than minor injury or PDO.

Recommendations:

- Conduct separate Table C analyses for (1) PDO and minor injury collisions and (2) fatal and severe injury collisions. This should be preceded by analyses of specific locations to confirm whether this split is in fact optimal.

3.2. ANALYSES BY DIFFERENT COLLISION TYPES

TABLE C:

Combines all types of collisions in the same analysis.

Observations:

- Different types of collisions have dramatically different distributions (e.g., run off the road collisions vs. rear end collisions).
- Caltrans of course already has some programs for identifying HCCLs for specific types of collisions (e.g., run-off-road collisions, wet weather collisions).

Recommendations:

- Conduct yearly analyses of specific types of collisions, especially those which (i) are fairly high in number and (ii) are likely to have a unique distribution. Examples would be pedestrian collisions, alcohol-involved collisions, collisions involving teenagers, etc.

4. CRITERIA FOR SELECTION OF LOCATIONS

TABLE C:

Above 99.5% Confidence Interval around a formula describing the relationship between volume and number of collisions. The formula has a feature for adjusted the rate (number per unit of volume) based on volume, but using that adjustment is set to '0,' i.e., the relationship between rate and volume is assumed to be a constant.

Observations:

- It is almost universally acknowledged that rate is not a constant over changes in traffic volume. We have confirmed a non-linear relationship between rate and volume within a number of Rate Groups.
- For intersections, ramps, and highway segments a function showing the relationship between volume (and other factors) and number of collisions (i.e., Safety Performance Functions [SPF]) combined with the Empirical Bayes method show great promise for improving estimates of expected collisions. One issue per the use of SPFs for highway segments is spatial correlation of collision clusters, especially along freeway segments.
- For highway segments, a method called Continuous Risk Profile (CRP) shows promise in determining high collision sites.

Recommendations:

- Discontinue use of the current Table C formula used to calculate the expected number of collisions.
- For intersections, ramps, and highway segments test the use of SPFs and the EB method.
- For highway segments test the use of the CRP method.

5. FORMAT AND CONTENT FOR REPORTING SITES

5.1. INFORMATION PROVIDED

(E.G., HIGHWAY FACTORS, NON-HIGHWAY FACTORS, COLLISIONS FACTORS)

TABLE C:

The Table C report provides location, Rate Group, total collisions in different time intervals, ADT, and number of fatal and injury collisions.

Observations:

- Some states (e.g., Colorado) provide a much richer set of information about HCCL sites.
- Much more information is available in the TASAS than is now provided in the Table C report.

Recommendations:

- Expand the Table C report to include:
 - Information already provided
 - Collision patterns
 - Comparison of collision patterns to other similar sites (e.g., within the same Rate Group)
 - Provide trends over time at the site compared to overall trends at similar sites.
 - Other information that could be derived from TASAS or could otherwise be linked to the type of site and collision pattern

5.1. INTEGRATED DATA SYSTEM

TABLE C:

The Table C report provides fairly limited data in a list format.

Observations:

- Table C appears to be distributed as a somewhat isolated report, i.e., apparently with no systematic link to the other data or to follow-up action.
- Providing Table C reports within the context of a broader data system may facilitate use and provide tracking capability.

Recommendations:

- Develop an integrated data system within which the Table C report is generated.
- The integrated data system would include:
 - Maps of Table C locations
 - Information on collision patterns available by pointing and clicking on a site
 - Tracking information including (i) results of investigation, (ii) installation of countermeasures, (iii) evaluation [i.e., pre-post collision history]

6. DATA QUALITY

Table C makes use of the TASAS (Traffic Accident Surveillance and Analysis System) database, which provides information about the California State Highway network. Variables in this system are important for identifying HCCL. The variables area are described in the Appendix (Transportation System Network [TSN], TSAR Reference Card).

There are three primary types of data

- 1 Highway Inventory
- 2 Volume Data
- 3 Collision Data

6.1. HIGHWAY INFRASTRUCTURE

TABLE C:

The State Highway System (SHS) includes more than 15,000 miles of highways, 14,000 ramps and 18,000 intersections. Variables include characteristics of the different types of sites. There are four types of Highway Inventory variables:

- Standard fields (functional classification, highway group, etc.)
- Highway fields (lanes and other design features)
- Intersection fields (configuration, traffic control device, etc.)
- Ramps fields (configuration)

Observations:

- Relatively minor issues include missing design information, overlapping sites (intersections within 250 feet of one another) and double listings. These may well be accounted for in Table C programming
- A more important issue is the small number of sites for some rate groups (see above)
- Important types of information are not included, such as (for highway segments), curvature and slopes. Sites within Rate Groups with features such as sharp curves and slopes will tend to have higher collision frequencies than other sites within the same Rate Groups.

Recommendations:

- Conduct a systematic audit of missing information, overlapping sites, etc.
- Develop a process for systematic screening for data issues.
- Consolidate Rate Groups with a small number of sites
- Add additional variables to TASAS that are known to affect collision frequencies.

6.2. TRAFFIC PATTERNS (INCLUDING VOLUME)

TABLE C:

Traffic volume data are obtained from Traffic Data Office (in Traffic Operations). An Average Annual Daily Traffic (AADT) is available for all intersections, ramps, and roadway segments. The calculation of Annual Average Daily Traffic (AADT) is performed once each year based on data collected during a year beginning October 1 through September 30. Volume is collected at all sites on a rotating basis once every three years.

Observations:

- Data are often out of date, many data points interpolated, etc.
- Some missing or out-of-range values
- Possible bias in volume estimates due to limited sampling

Recommendations:

- Determine impact of current sampling scheme on volume estimates
- Increase number of counts
- Test models of extrapolation and interpolation (current methods appear inadequate)

6.3. COLLISION DATA

TABLE C:

Collision data are obtained from the California Highway Patrol (CHP) from a database called SWITRS (Statewide Integrated Traffic Records System) SWITRS is intended to include all police-report traffic collisions in the state. Collision data are extracted by CHP from the SWITRS database and contain information about collision aspects and party involved, coded by CHP, as well as site location, coded by Caltrans. Between 1994 and 2004, more than 1,800,000 accidents were recorded on Californian State Highways.

Observations:

- Underreporting (based on numerous studies)
- Inaccurate information (internally inconsistent, out-of-range)
- Issues of linking collision data to location

Recommendations:

- Create programs to do systematic range and missing value checks
- Prepare reports on out of range and missing data as feedback to CHP and other police agencies.
- Test models of extrapolation and interpolation

7. APPROACHES OTHER THAN SITE-SPECIFIC APPROACHES

7.1. INDIVIDUAL SITES VS. TYPES OF SITES

TABLE C:

Focus on individual sites (intersections, ramps, or 0.2 mile segments)

Observations:

- Patterns of individual sites showing high collision concentrations make reflect design features that impact safety. Possible examples include access points on limited access HOV lanes, excess collisions on freeway lanes near ramps, etc.

Recommendations:

- Statistical models such as Safety Performance Functions, Empirical Bayes Methods, and the Continuous Risk Profile method, should be developed to identify patterns of collisions related to various design features. Examples of this are underway for HOV lanes and ramps.

7.2. CORRIDORS

TABLE C:

Focus on individual sites

Observations:

- In some cases HCCLs will be adjacent or near one another.
- In some cases a series of segments (or intersections) have a relatively low density such that a Table C HCCL would not be identified, however, amounting to a fairly high density of collisions if one were to view segments longer than 0.2 miles. An example would be a rural roadway with relatively high traffic and with a high cumulative number of collisions spread somewhat uniformly along an extended section of roadway.

Recommendations:

- Develop a statistical methodology for identifying corridors. Such a method could be developed by using a “sliding window” of different lengths. A method for looking at segments of different lengths (e.g., 1/2 mile segments) is being developed in the context of developing the 5% report for the Strategic Highway Safety Implementation Plan (SHSIP). Another approach is to examine traffic density in “natural” segments, i.e., segments between intersections or exchanges. Finally, another approach plots collisions using GIS and then uses existing software to identify clusters of collisions—such methods allow for clusters of varying lengths and densities. Each of these methods is feasible within the context of the current TASAS data system.

TABLE OF CONTENTS

1. Background	1
2. The current report	3
3. Physical Structure of Analysis	5
4. Temporal Structure of Analysis	12
5. Choice of Outcome(s)	14
6. Criteria for Selection of Locations	16
7. Format and Content for Reporting Sites	24
8. Data Quality	26
9. Approaches other than Site-Specific Approaches	31
Appendices	33

1. BACKGROUND

1.1. TABLE C

There are approximately 190,000 reported collisions on California state routes annually. One of the department's goals is to reduce the number and severity of these collisions. To help achieve this goal, every quarter the Department publishes a list, called "Table C," of high concentration collision locations (HCCL). There are 170 traffic safety investigators in Caltrans who review about 10,000 locations annually. Roughly 700 improvements are initiated annually as a result of the HCCL program. Traffic investigators also receive an annual "Wet Table C" that identifies high wet pavement collision concentration locations.

Table C makes use of the TASAS (Traffic Accident Surveillance and Analysis System) database, which provides information about the highway network, such as design characteristics and traffic volumes, as well as a full history of accidents during the past ten years. Across California, more than 15,000 miles of state highways, 14,000 ramps and 19,000 intersections are detailed in the Highway Database. Information is obtained and updated by reviewing construction plans and working jointly with district TASAS coordinators

The data are distributed into four different tables. The first three provide the description and design characteristics of the highway sites studied, which are classified as segments, intersections, and ramps (Highway Database). The fourth table provides detailed information about all accidents reported by the police during a period of about 10 years (Collision Database).

1.2. CALTRANS REVIEW OF TABLE C

In 2002 Caltrans completed a review of the HCCL investigation process, making the following short-term and long-term recommendations.¹

1.2.1 SHORT-TERM TABLE C RECOMMENDATIONS

1 IDENTIFY AND ELIMINATE REPEAT LOCATIONS

Repeat locations are defined as 100% the same postmile limits as any "required" location identified during the previous 3 quarters. Repeat locations will be screened out and will not be included in the list sent to the districts for investigations.

2 IDENTIFY AND ELIMINATE OVERLAP LOCATIONS

Overlap locations are defined as an overlapping segment of 51% to 99.99% with any "required" location identified during the previous three quarters. Overlap locations will be screened out and not sent to the districts.

3 COMBINE ADJACENT HIGHWAY LOCATIONS

These locations are defined as highway segments that are adjacent to one another. The adjacent locations will be combined in the report to the districts and will be done in a single investigation. Combined locations will not exceed one mile in length.

4 SEND OUT ONLY "REQUIRED" LOCATIONS

Only those locations marked with a "Req" will be sent to the districts.

5 UPDATE INTERSECTION TRAFFIC VOLUME

Update intersection traffic volume.

¹ Table C Task Force: Summary Report of Task Force's Findings and Recommendations.

1.2.2 LONG-TERM TABLE C RECOMMENDATIONS

1 MODIFY THE SELECTION CRITERIA

Minimum number of collisions and statistical significance threshold could be evaluated.

2 WEIGH THE SEVERITY OF COLLISIONS

Fatal, injury, property damage collisions only. Should there be a prioritization for investigations by placing a weighted factor on collisions based on severity?

3 ANALYZE THE SEGMENT BY COLLISION OR REVISE LENGTH

Should the selection of location be made on the location of collisions and/or collision rate and not constrained by the segment length of 0.2 mile?

From this review, and in light of the long-term recommendations, Caltrans initiated Task Order 5215 with the California Partners for Advanced Transit and Highways (PATH) and the University of California, Berkeley Traffic Safety Center (TSC). PATH and TSC proposed to evaluate the methodologies used for the identification of high-concentration collision locations.

2. THE CURRENT REPORT

2.1. BASIC TASK

The primary mandate of the current project is to evaluate current methodologies used to create Table C. The methods included:

- 1 Evaluation of methodologies used by different states
- 2 Conducting a detailed evaluation of elements of the Table C method, including those mentioned in the long-term recommendations above
- 3 Making recommendations for modification of Table C

2.2. STEPS ACCOMPLISHED IN PREPARING THIS REPORT

- Extensive review of literature and interviews with other state DOTs (literature review completed August 2005)
- Extensive consultation with national experts in this area (throughout the project period)
- Sample data analyses using TASAS data (report completed May 2006)
- Extensive consultation with Caltrans safety personnel (throughout the project period)

2.3. STRUCTURE AND ORGANIZATION

Approaches to identifying high collision concentration locations (HCCL) can be defined in terms of six basic issues, as follows:

- 1 PHYSICAL STRUCTURE OF ANALYSIS UNITS**
 - Should analyses be conducted independently within rate groups, or should all categories of sites be compared together?
 - If analyses are to be conducted rate groups, how should rate groups be defined?
 - For analyses of roadway segments, how should such segments be subdivided in the analysis?
- 2 TEMPORAL STRUCTURE OF ANALYSIS**
 - Length of time used to calculate the base
 - Length of time used to estimate the risk at a specific site
 - Frequency with which the analysis is conducted (e.g., quarterly, biannually, yearly?)
- 3 CHOICE OF OUTCOME(S)**
 - Type of collision (PDO, injury, fatality)
 - Weighting of different types of collisions
 - Units (e.g., number, cost)
 - Denominator (unit, distance, VMT).
- 4 CRITERIA FOR SELECTION OF LOCATIONS**
 - Absolute number, relation to a distribution
 - Categorical vs. graded
 - Level of modeling (e.g., use of average, PDF, or Bayesian approach)

FORMAT AND CONTENT FOR REPORTING SITES

- Information provided (e.g., highway factors, non-highway factors, collisions factors)
- Level of analysis

DATA QUALITY

- Highway infrastructure
- Traffic patterns (including volume)
- Collision data

2.4. BASIC PRINCIPLE

The method for determining HCCL will determine which locations are chosen. Any one of a number of decisions will result in a different—sometimes very different—set of chosen locations. A basic guiding principle throughout this report is that benefit per unit of cost should be maximized. This principle has been articulated by Ezra Hauer as follows: "...money should go where it achieves the greatest effect in terms of saving accidents and reducing their severity." To not follow this principle would mean that it is justified "...to save one accident when, for the same money, more could be saved. Such justifications are not easy to find."²

To determine whether a site or set of sites will yield the "biggest bang for the buck" can only be accurately determined after an on-site investigation. Such an assessment requires an understanding of the characteristics of crashes (**a measure of the impact of collisions**) at a site and then an estimate of the effectiveness and cost (in the case of Benefit/Cost) of the countermeasure. The former (i.e. the knowledge about the crashes) can be determined with some accuracy prior to an onsite investigation. The latter (i.e. the countermeasure benefits and costs) can only be exactly determined after a site-specific investigation. However, each decision made about selecting HCCLs will have an impact on eventual benefit cost ratio, and the aim is anticipate this as accurately as possible in the screening phase.

In the following section we will evaluate in detail the issues listed above. In each case, the guiding principle will be the likelihood that a particular decision will lead to most effective use of highway safety resources.

Table C is a method for screening sites. As such, it is just one of several steps in a process to identify sites with the greatest potential for improvement. Other steps include site investigation (diagnosis), countermeasure selection, and prioritization.

The ultimate goal of this process is to choose sites with the most potential for improvement. At this stage of screening we cannot know exactly which sites will ultimately have the largest potential for improvement. Different methods and approaches will generate different—even very different—sites. Different sites may have much different potential for improvement. Our intent is to choose methods/approaches that will increase the likelihood that sites chosen will have the greatest potential for improvement.

² Hauer, E. Screening the road network for sites with promise, TRB, 2002.

3. PHYSICAL STRUCTURE OF ANALYSIS UNITS—WHAT IS A SITE

3.1. PHYSICAL STRUCTURE OF ANALYSIS

By definition, the process of identifying HCCLs depends on being able to identify specific locations. In California, the State Highway System is divided into three major groups: Intersections, Ramps, and Roadway Segments. Each of these major groups is further divided into subgroups (or “Rate Groups”) based on various dimensions. All analyses are conducted independently within Rate Groups.

The main issues:

- 1 Should analyses be conducted independently within Rate Groups, or should all categories of sites be compared together?
- 2 If analyses are to be conducted Rate Groups, how should Rate Groups be defined?
- 3 For analyses of roadway segments, how should such segments be subdivided in the analysis?

3.1.1. SHOULD ANALYSES BE CONDUCTED WITHIN CATEGORIES OF LOCATIONS OR SHOULD ALL LOCATIONS BE COMPARED TOGETHER?

The current Table C procedure for selecting sites involves comparing individual sites to the average of all sites within that particular Rate Group. Taking intersections as an example, a “base rate” for each Rate Group is derived by calculating the number of collisions per 1 million vehicles for the entire set of intersections within the Rate Group. The expected number of collisions for a particular intersection is then calculated by multiplying the base rate by the traffic volume at that intersection. If the actual number exceeds the expected number by a significant amount (see section below on statistical modeling and statistical tests), then the intersection is considered to be an HCCL. For any particular ADT this method maximizes rate, since, at that volume the number of collisions for achieving significance will be reached only if the rate is substantially higher than the base rate for that Rate Group.

However, base rates vary substantially among different rate groups. One consequence is that, for a particular volume, selected sites identified as HCCL in rate groups with low base rates may have much lower rates, and of course lower collision frequencies, than sites not selected in rate groups with high base rates. An example is intersections in the “No Control” category, which are subdivided into rural, suburban and urban. The base rates are 0.11, 0.35, and 0.06, respectively, for rural, suburban, and urban. In this case, intersections selected in the urban and rural categories are likely to have much lower rates than intersections selected in the suburban category; this means, of course, that many suburban intersections with relatively high rates (compared to No Control intersections in rural and urban areas) will not be selected as HCCLs. There are some intersections in rural areas that are not chosen that have higher risk than the criteria for urban intersections. There are some intersections in urban areas that will be selected that would not be selected if in rural areas. Finally, the overall level of risk of selected sites will be lower when selection is done separately for urban and rural intersections.

The same issue arises when other comparisons are made among other rate groups for intersections and among rates groups for ramps and highway groups (see Appendix A where the list of rate groups in the current Table C method is given).

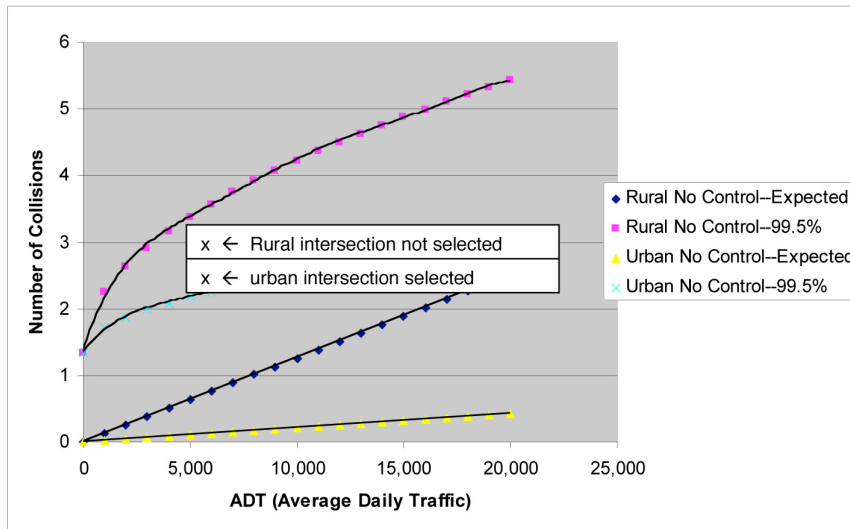
³ The phenomenon can be illustrated in the following way. Suppose that we are asked to put together the best baseball team comprised of members of professional baseball teams in California. Suppose further that we are asked to chose half the players from Major League teams, and the other half from Minor League Teams. This would be optimizing locally (within Major and Minor teams), but certainly would not be optimizing globally (i.e., producing the best possible baseline team).

Table 1

RATE GROUP	BASE RATE	+ ADT FACTOR	PC T	PC T	PC T	INTERSECTION TYPE *	CONTROL	TYPE	AREA	ACC COSTS (\$1000)	
										F+I	ALL
I 01	0.11	0.0000	3.5	42.0	45.5	F, M AND S	NO CONTROL		RURAL	371.4	171.2
I 06	0.35	0.0000	0.8	32.3	33.1	F, M AND S	NO CONTROL		SUBURBAN	147.9	51.6
I 11	0.06	0.0000	2.6	42.8	45.4	F, M AND S	NO CONTROL		URBAN	249.7	115.5

The approach used in the Table C method, i.e., conducting analyses within categories of roadways, can generally be described as “maximizing locally” instead of “maximizing globally.” Unless risk (however defined) is spread evenly across rate groups, maximizing locally will inherently result in a sub-optimal global maximum.³

Figure 1



It might be argued that global optimization is preferable because it produces the segments that have the highest overall risk (however defined).⁴ However, there are several types of factors that suggest maintaining some levels of categorization when evaluating road risk.

First, risk across some categories may not be inherently comparable, at least not while using current approaches. For example, risk in intersections is defined in a different way (per entering vehicles) than risk along a roadway segment (per roadway mile or per vehicle mile traveled).

Second, constraints brought about by political considerations or funding streams may dictate that risk be evaluated within categories defined by particular categories. One example may be the rural-suburban-urban distinction presently embodied in the Table C rate group structure.

Third, and perhaps most importantly, the cost and effectiveness of countermeasures may not be equal across different categories. For example, at intersections with lower rates (or lower frequencies), cost of countermeasures may be lower, or effectiveness may be higher, which would tend to increase the benefit cost ratio for these intersections in relation to intersections with higher rates (or higher frequencies). This is often claimed anecdotally, but there does not appear to be any specific information available on this topic. This is a topic for further research.

⁴ Just as choosing a baseball team from among all professional players in California in a combined group will result in the best team.

Overall, there appear to be reasons to move in the direction of global optimization, but to maintain a rate structure since it serves a purpose, for example, when one or more of the reasons given above apply.

TENTATIVE RECOMMENDATIONS

- 1** Study the implication of local optimization (determine the extent to which local optimization reduces global optimization). This can be done by comparing **(1)** the rates and number of collisions identified as Table C HCCLs to those that would be chosen within the group of sites as a whole and **(2)** comparing the cost and effectiveness of treatments within different rate groups
- 2** Consider separately the relevance of each dimension that is used, or could be used, to define rates groups. Each dimension defining rate groups should be justified in terms of one or more of the reasons given above.

3.1.2. IF ANALYSES ARE TO BE CONDUCTED USING RATE GROUPS, HOW SHOULD RATE GROUPS BE DEFINED

As stated in the previous section, California's State Highway System is divided into three major groups: intersections, ramps, and roadway segments, and each group is defined further into subgroups called "Rate Groups." Virtually every approach that we reviewed divides the roadway into categories in one way or another. The basic rationale in every case is to define groups with common characteristics and then conduct a comparison within these groups. A site with a higher risk, however defined, with respect to other similar sites is selected as a candidate for further investigation. The informal rationale often given is that this it is necessary to compare within similar categories "apples to apples, and oranges to oranges."

The variables used to differentiate rate groups within the broad categories of intersection, ramp, and roadway are as follows:

Formally, the approach defines two types of site characteristics:

SET A: SITE CHARACTERISTICS THAT ARE USED TO DEFINE CATEGORIES.

In Table C, the most important such characteristic is of site (intersection, ramp, roadway).

Intersection

- Control Type (no control, stop and yield [except 4-way])
- Intersection Type (F,M,S versus T, Y, Z)
- Area (rural, urban, suburban)

Ramp

- Ramp Type (frontage road, etc.)
- Ramp Area (1-4, 1-3, etc.)
- Area (rural, urban, suburban)

Roadway

- Highway Type (conventional two lanes or less, etc.)
- Terrain or ADT (flat, etc.)
- Area (rural, urban, suburban)

SET B: ALL OTHER CHARACTERISTICS THAT COULD AFFECT COLLISIONS

This includes any characteristics which are not part of Set A. Some of these characteristics are variables available in TASAS (volume, shoulder width, speed limit, etc.) and—of great importance—some are not (curvature, slope, etc.).

The method utilized holds constant characteristics within Set A, and looks for variation in collisions within categories defined by Set A that are presumably caused by some characteristics in Set B, and do not arise simply by chance. It is presumed that excessive collisions, however, arise either by chance or are defined by some characteristic in Set B.

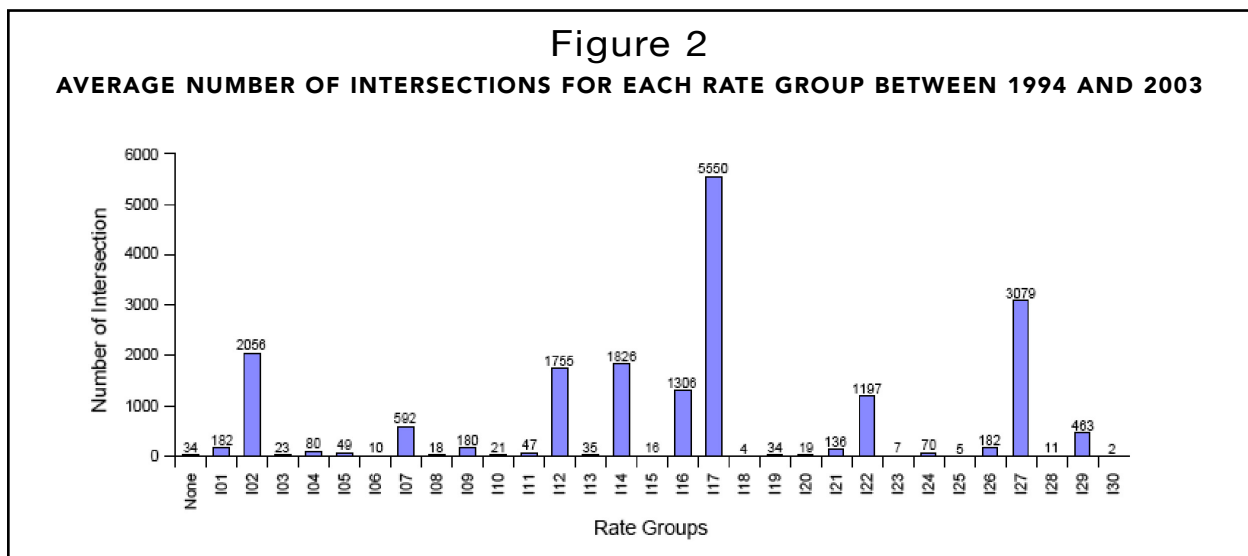
The main task in this section is examining the rationale for defining characteristics in Set A versus those in Set B.

PRINCIPLE 1: Exclude from Set A characteristics that are often used to defined countermeasures.

We don't want Set A to include characteristics that would often be identified as countermeasures. One example is rumble strips. Using this characteristic to define Table C categories might mean that it would be missed as a possible factor (when absent) in run off the road collisions, and therefore might not be considered as a countermeasure. We generally would want Set A to consist of categories that are not amenable to change.

PRINCIPLE 2: Include in Set A characteristics that define fundamental differences in the nature of sites.

We want Set A to include characteristics that define a basic or fundamental difference in type of site. Intersections, ramps, and roadways are very different entities. Intersections and ramps are usually discrete entities whereas roadway segments are of variable length. Risk is defined in various ways. For example, risk by usage is defined in different ways: **(1)** risk in intersections is defined as the number of collisions divided by the sum of the number of entering vehicles; **(2)** risk at ramps is defined as the number of collision divided by the number of vehicles passing through the ramp; and **(3)** risk on roadway segments is defined as the number of vehicle miles traveled (VMT). Risk can also be defined independent of use: **(1)** risk for intersections can be defined simply as the frequency; **(2)** risk for ramps and roadways can be defined in terms of the number of collisions per unit of length (density). There appear to be fundamental differences in how risk is defined in these three major site categories. Some of these same considerations might apply to other divisions defining rate groups, for example, signalized versus unsignalized intersections, two lane roadways versus freeways, etc. These considerations suggest that these dimensions be maintained.



Cost and effectiveness of countermeasures may not be equal across different categories. For example, at intersections with lower rates or frequencies, cost of countermeasures may be lower, or effectiveness may be higher, which would tend to increase the benefit cost ratio for these intersections in relation to intersections with higher rates or frequencies.

PRINCIPLE 3: We want Set A characteristics to define categories that are of sufficient size that statistical analyses are meaningful.

Clearly, categories that are too small lead to highly uncertain estimates of risk. Taking intersections as an example, we have noted that some of the rate groups have very small samples. Second, there is a very uneven number of sites across categories, leading to substantial differences in statistical variation and therefore, especially among categories with few sites, increased false negatives and false positives.

PRINCIPLE 4: Political or funding constraints.

Such constraints might operate along different dimensions. For example, whether the location is rural, urban, or suburban is a major dimension defining Table C rate groups, and funding mechanisms may differentiate these categories. The same may be true for type of highway, such as conventional two lane highway versus freeway.

In addition, constraints due to political considerations or funding streams may dictate that risk be evaluated within categories defined by particular categories. One example may be the rural/suburban/urban distinction presently embodied in the Table C rate group structure.

In the previous section we discussed the concept of dividing the roadway into categories of sites and concluded that a rationale should be to provide for each characteristic that defines the categories. In this section we examine the specific rate groups used for Table C and address the potential rationale for each. In the following section we will outline the considerations and their application to specific characteristics that define, or could be used to define, rate groups:

- Review variables that define rate groups and determine which, if any, can be eliminated. Maintain categories that meet basic criteria above. One possibility would be to combine rural, suburban, and urban categories while maintained.
- Review variables that currently do not define rate groups and determine which, if any, should be added.
- Examine differences in outcome across different rate groups.
- Develop formal rationale for roadway categories based on similarity in type of traffic flow and collision patterns.
- Attempt to equalize, at least approximately, the number of sites in each category—this could be accomplished by either combining similar but smaller categories, or splitting larger categories when appropriate.
- Determine if the assumption about countermeasures (CM) effectiveness stated above is valid.

3.1.3. SEGMENTATION WITHIN CATEGORIES (FIXED WINDOW, MOVING FIXED WINDOW, VARIABLE WINDOW, CONTINUOUS)

Ezra Hauer has discussed this issue at length in several publications and this issue has also been discussed in the *SafetyAnalyst* White Paper on Network Screening (2002).

Ezra Hauer and others have discussed pros and cons of several methods:

1 ENTIRE ROAD SECTION

One possible hose is an entire road section. This entails averaging over the entire road section. In Table C road sections can be of varying length, from a fraction of a mile to several miles in length. For long road sections peaks in collision risk will be washed out by averaging with lower collision strengths. Shorter road sections, while not having the advantage of mixing wide variations in risk, will be much more unstable, and false positives are likely to arise.

2 SEGMENTS OF FIXED LENGTH

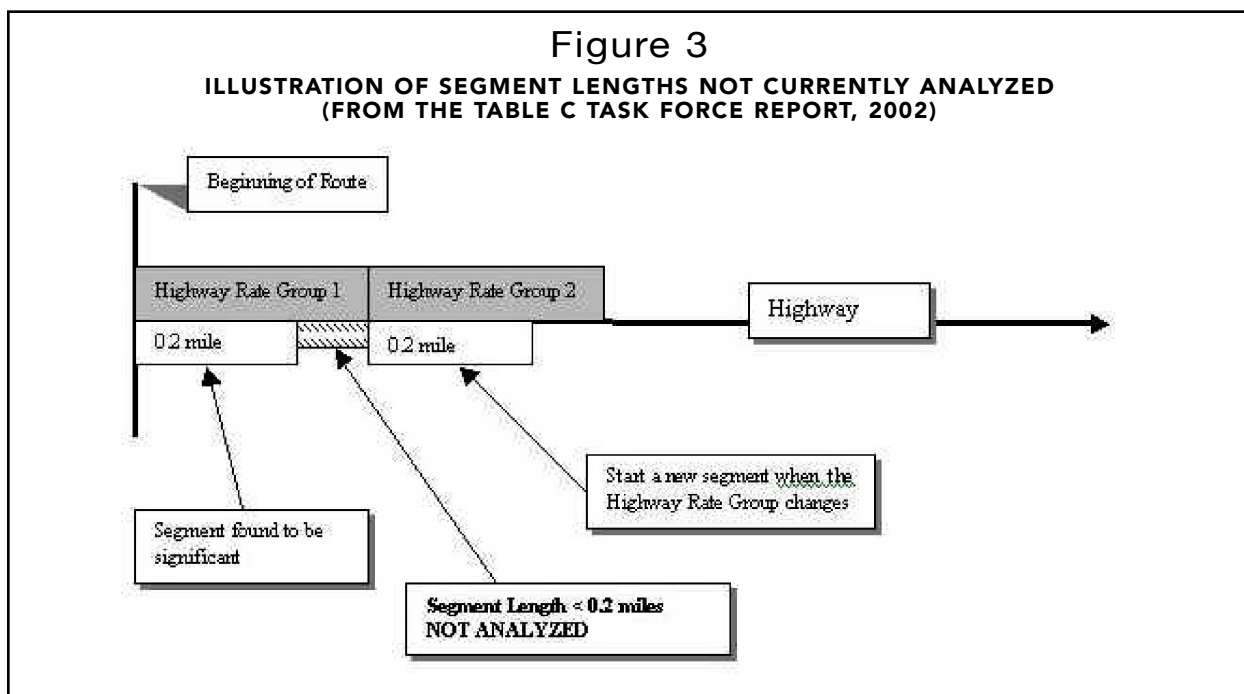
Description

Pros and Cons

3 PEAK OF VARIABLE LENGTH

Description

Pros and Cons



⁵ Table C Task Force Report, 2002.

4 TABLE C MOVING WINDOW APPROACH

Table C currently uses a fixed-length moving window approach, moving a frame of 0.2 mile which moves in increments of 0.02 miles. With each increment a statistical test is performed and 0.2 mile segments that are in the top 0.5% region are selected for detailed study (see xx for more detail).

There are two important concerns with the fixed-length moving window approach as utilized in producing Table C.

The first concern is that since the window is fixed at 0.2 miles, some segments will not be evaluated. This can happen in two ways. First, a highway segment with a length of less than 0.2 miles will not be evaluated. Second, there can also be “left-over” segments when a 0.2 mile segment is found to be significant and the remaining portion of the entire segment is less than 0.2 miles. This concern was noted in the Table C Task Force Report⁵: “The Table C program does not analyze highway segments less than 0.2 miles in length. Examples include segments just before intersections, route breaks and district boundaries, and at changes in rate group (Figure 3).”

The second concern is that the fixed window may not “fit” actual risk profiles. The segment of roadway with increased risk may be shorter or longer than the length of the fixed window or may be of variable magnitude. As we have argued in our paper describing the Continuous Risk Profile (CRP), both false negatives and false positives can arise. The CRP method addresses this concern by allowing a much closer “fit” to the underlying risk instead of forcing an arbitrary 0.2 miles (or any other fixed length). More discussions of CRP can be found in the next section or in the referenced publication (XX).

The state of Colorado has a very different approach in screening sites for potential safety improvements. (References XX) For example, when evaluating safety risks on interstate freeways a segment is defined as a section of the freeway between junctions or entry and exit ramps. For other typical roadways, instead of using a segment of fixed length, a segment is defined as a stretch between two intersections or junctions.

3.2. CONCLUSION

Within the roadway segment category, the choice of 0.2 mile segment is somewhat arbitrary. Based on the review of historical collision data, many high-risk locations are usually smaller in size. The use of 0.2 mile segment may mask the safety risk levels and thus causes a miss of the high-risk locations. In addition, by using a fixed-length segmentation, artificial limitations are imposed on the system when segments smaller than the fixed size were not included automatically in the process.

To summarize, we think that the CRP may be useful as part of the Table C method that identifies risk on highway segments. We suggest testing the CRP as a possible alternative to the moving fixed window approach for the portion of the Table C method that analyzes highway segments, as opposed to intersections and ramps.

4. TEMPORAL STRUCTURE OF ANALYSIS

This section deals with the time frames or windows, from which the historical data are relied on, and for which the analysis of safety risks are estimated, and then by which the frequency of outputs that are generated in the methods for identifying HCCLs.

In the process of identifying HCCLs and generating Table C is, the amount of data that are used for establishing the baseline or expected numbers dictate the “thresholds” that are the most critical variable in the statistical analysis. In addition, there are various complications when a particular time window is selected for screening the high-risk locations as the stability and consistency of data vary due to the nature of fluctuations in collision numbers and the “reversion to the mean” phenomenon that is of great importance in identifying the outliers of a distribution by statistical analysis. Moreover, the frequency of outputs or reports of Table C or other HCCL screening methods will also have significant impacts on the reliability and accuracy of HCCL, as well as the the efficiency of resource utilization needed for the follow-up safety investigations.

In summary, the main issues within the temporal structure of HCCL screening are:

- 1 Length of time used to calculate the base
- 2 Length of time used to calculate the risk at a specific site
- 3 Frequency with which the analysis is conducted

The major observations and recommendations for these issues are provided in the sub-sections below.

4.1. LENGTH OF TIME USED TO CALCULATE THE BASE

TABLE C:

Based rates are calculated using three years of data.

Observations:

- If a Rate Group has a sufficient number of member sites, three years ought to be sufficient to provide stable estimates of Base Rates; the variation in rates becomes fairly small with as few as 20 member sites in a Rate Group.

Recommendations:

- Maintain a three year period.
- Evaluate trends over an extended period of time to determine if there is “drift” in underlying Base Rates.

4.2. LENGTH OF TIME USED TO ESTIMATE THE RISK AT A SPECIFIC SITE

TABLE C:

Tests are conducted using 3 months, 6 months, 9 months, 1 year, 2 years, and 3 years.

Observations:

- Any time period less than a year is too short to a stable estimate of HCCL, no matter how calculated. One year is adequate for sites with high volume but may not be adequate for sites with low volume.
- The current method used (and most other methods) are aimed at determining elevations in fixed risk at particular sites (i.e., assume that risk is constant over time). The method (and most other methods) is not designed to detect changes in risk over time.

Recommendations:

- Eliminate estimates based on any time period less than one year.
- For all sites, use a method proposed by Ezra Hauer to determine stability of estimates. Use that method to decide whether, for particular sites, the time period should be one year, two years, or three years.
- Utilize a method recommended in SafetyAnalyst for determining changes in risk. This should be routinely applied to all sites on a quarterly basis, and especially in sites that are experiencing other changes.

4.3. FREQUENCY WITH WHICH THE ANALYSIS IS CONDUCTED

(E.G., QUARTERLY, BIANNUALLY, YEARLY?)

TABLE C:

Quarterly report.

Observations:

- The survey conducted by the Table C Task Force indicated that Caltrans users of the Table C report are in favor of having a quarterly report (as opposed to biannually or yearly).

Recommendations:

- The Table C report should be produced quarterly.
- However, a standard analyses for HCCLs should be done only on a yearly basis.
- Other quarters should include reports on sub-topics, especially analyses of potential change in risk (see above).

5. CHOICE OF OUTCOME(S)

This section covers the discussions of crash types and their severity levels in the identification of HCCL. The severity levels of crashes, when incorporated by weighting factors in the selection function, have significant effects on data stability and the outcome of HCCL lists. In addition, because different types of crashes may have different distribution patterns, an alternative in HCCL analyses is to conduct separate screening and processing for various crashes.

5.1. WEIGHTING BY LEVEL OF SEVERITY (PDO, INJURY, FATALITY)

TABLE C:

Caltrans The current method in Table C currently does not weight collisions based on severity in determining high collision concentration locations.

Observations:

Many approaches to injury severity weighting use variations on the “equivalent property-damage-only” (EPDO) method. In this method, weights of fatal and injury crashes are compared to the weight of a PDO collisions. For example, the state of Iowa currently weights PDO collisions by 1, injury collisions by 5, and fatal collisions by 8 (2). Researchers at the University of Limburgh suggest weights 1, 3, and 5, respectively. Another approach to using weights is to use numbers that reflect the actual cost of each collision. Whether EPDO or cost approach is used, the ratio of the weights is usually based on average total costs of property, injury, and fatality collisions. Using these weights, a severity index is developed for each highway segment using the following formula:

$$SI = [W_f F + W_m M + W_c C + P] / T$$

Where:

SI is severity index

W_x are weights for fatal, major, and complaint of pain collisions

P is PDO collisions

T is total crashes at site (2)

Highway segments can be ranked by severity index, or the severity index of other criteria such as the crash frequency, crash rate, or can be integrated as part of the quality control methods discussed previously (2).

Each of these approaches has two major flaws:

- 1 If fatality is weighted too heavily it creates instability in the estimates, since fatality is rare
- 2 It assumes that collisions of different severity are similarly distributed across locations. In fact, PDO, injury, and fatal collisions have substantially different distributions. We have compared the relative distribution of fatalities, severe injuries, and minor injuries. It appears that major injuries are more closely related to fatalities than to minor injuries.

5.1.1. DISCUSSION:

It is quite common to weight by injury severity, although the current Table C methodology does not do so. The primary reason to weight collisions is to account for the increased burden or cost of specific types of collisions. For example, putting a larger weight on fatalities will mean that locations with fatal collision are more likely to be identified as high risk locations.

However, there are several issues. One issue arises when of severe collisions are more heavily weighted. Severe collisions tend to be rarer, and therefore the stability of estimates will be reduced, i.e., some locations might be

identified based on one or two fatalities that arose “by chance” at those particular locations, and not because of something inherent in the locations. Clearly, in weighting by severity there is a trade-off with statistical stability.

Another issue is how to determine the proper and equitable weighting. Weighting by severity of injury is the approach used most often. However, other factors might be used in weighting, such as the cost of congestion or delay, which may be high even in PDO collisions.

A third issue is the relevance of the weighting by severity to highway factors. Severity does not always result from, nor is it sensitive to, highway factors, since severity depends on many other factors such as such as vehicle speed, vehicle type, seat-belt use, and other non-highway characteristics (2).

Recommendations:

Conduct separate analyses for (1) PDO and minor injury collisions and (2) fatal and severe injury collisions. This should be preceded by a study to confirm whether this split is in fact optimal.

5.2. ANALYSES BY DIFFERENT COLLISION TYPES

TABLE C:

The current method in Table C cCombines all types of collisions in the same analysis.

Observations:

Different types of collisions have dramatically different distributions (e.g., run off the road collisions vs. rear end collisions), although. Caltrans of course already has some programs for identifying HCCLs for specific types of collisions (e.g., run-off-road collisions)

There are three types of approaches to providing more information in HCCL reports.

- 1 Create a “Table C” for specific kinds of collisions. This approach is already used for “wet” highway collisions in order to generate a Wet Table C. The goal of this approach is to help engineers identify where slippery pavements might be the cause of an unusually high number of collisions. If desired, similar tables could be created such as a Roll-Over Table C, Broadside Table C, Rear-End Table C, a DUI Table C, etc.

Recommendations:

- Conduct yearly analyses of specific types of collisions, especially those which (i) are fairly high in number and (ii) are likely to have a unique distribution. Examples would be pedestrian collisions, alcohol-involved collisions, collisions involving teenagers, etc.

Conduct yearly analyses of specific types of collisions, especially those which (i) are fairly high in numbers and (ii) are likely to have a unique distribution. Examples would be pedestrian collision, alcohol-involved collisions, collision involving teenagers, etc.

6. CRITERIA FOR SELECTION OF LOCATIONS

6.1. METHOD FOR CHOOSING HCCLs

Most methods for choosing HCCLs begin by calculating an expected number of collisions for particular sites and determining the distribution around the expected number. Then, the actual number of collisions is determined for individual sites and a site is designated as an HCCL if the actual number exceeds the expected number by a certain amount (for example, if the number is above the 95% confidence interval). Four such methods have been reviewed: the method used for producing Table C (N_E), the Safety Performance Function (SPF), the Empirical Bayes (EB) method, and a newly developed method called the Continuous Risk Profile (CRP). The first three of these methods rank sites based on their position in an expected distribution and then select HCCL sites that are on the upper end of that distribution. Each of these methods has been applied to both discrete sites (sites without a distance dimension such as intersections and ramps) and extended sites (such as roadway segments). The CRP, applied so far only to roadway segments, calculates a base density of collisions (such as number per unit of distance) and then produces a continuous density profile in relation to the base density. Road segments of variable length with high profiles can then be chosen as HCCLs.

In the following, these four methods will be compared in the task of choosing HCCLs.

- Table C (N_E)
- Safety Performance Function (SPF)
- Empirical Bayes (EB)
- Continuous Risk Profile (CRP) (for highway segments only)

6.1.1. TABLE C METHOD

For the Table C approach the expected number is calculated by the following formula:

The average number of accidents

$$(1) \quad N_E = ADT \times t \times L \times R_E \div 10^6$$

Where:

ADT = Average Daily Traffic, vehicle per day

t = time, in days = #quarters x days/quarter (Table C)
x days/time period (Table B)

L = length, in miles

(= 1 for Ramps and Intersections)

R_E = Average Accident Rate, in accident/million vehicle (ACCS/MV) or accident/million vehicle mile (ACCS/MVM)
= Base Rate + ADT factor

Based on the type of facility, each type of highway, ramp or intersection is placed in a Rate Group. Each Rate Group has Base Rate and ADT factor that are determined by looking at all accidents in a three year time period. (See Appendix B, C, &D for the Rate Group of Intersection, Ramp, and Highway).

Then, a 99.5% upper confidence interval is calculated as follows:

$$(2) \quad N_E + 2.576(N_E)^{1/2} + 1.329$$

N_E is defined for each site. If the actual number at that site is greater than the 99.5% confidence limit, then the site is designated as an HCCL. The concern in this section is whether N_E is a good estimator of the number of collisions that will occur at a particular site over a period of time. N_E is relatively easy to calculate and understand. However, there are three several primary limitations.

First, for most of the rate groups (all of the intersection and ramp segments, and most of the highway segments) the ADT factor is set to '0,' so that the rate is not adjusted by the ADT factor. This means the rate is assumed to be constant over volume and therefore that the number of collisions is a linear function of volume. Virtually all researchers now working in this area maintain that the rate of collisions is not constant over volume, and, equivalently, that the number of collisions is not a linear function of volume (xx). Several empirical checks for specific types of sites in TASAS have shown that the rate changes with volume. Depending on the actual relationship between rate and volume, the implication of assuming a linear relationship is that both false positives (at sites with high ADT) and false negatives (at sites with low ADT) will be increased.

Second, the ADT factor adds an amount proportional to the ADT that is added to the rate. This is not a standard statistical approach to accounting for ADT. The implication is that there is no accepted statistical method available for estimating this parameter, other than checking manually to see how it fits the data.

Third, N_E does not permit including variables other than traffic volume, such as shoulder width, number of lanes, etc. The implication is that this can result in biased estimate of expected frequency and increase its variance. However, it should be noted that the implicit consideration in the Table C approach takes into account some roadway attributes by categorizing roadways into many different rate groups.

Fourth, and applying only to roadway segments, the method does not account for potential serial correlation among highway segments. Collision numbers are serially correlated in adjacent sites because hot spots will tend to generate secondary collisions in the neighboring sites. The implication is that this will affect the estimate of variability and therefore the confidence interval calculation.

Fifth, the method implicitly assumes that all the factors causing the high collision rates in the segment reside within. When the collisions rates are high due to the secondary collision in the vicinity, this method will also detect the neighboring sites without showing the relationship between its collision rates with the adjacent sites. This will result in detecting multiple sites that are adjacent to each other. (Note that this was one of the issues addressed by Safety Engineering during the survey conducted in 2003)

IN SUMMARY, THE TABLE C APPROACH POSSESSES THE FOLLOWING CHARACTERISTICS:

Strengths:

- Relatively easy to calculate and understand
- Allows variation in collision frequency as a function of traffic volume
- Allows a non-linear relationship between number of collisions and traffic volume (although via a functional relationship that does not lend itself to modeling the non-linearity)

Weaknesses:

- Biased if assumption about constant rate is not true (in the case where rate actually declines with volume, false positives will arise at sites with low volume, and false negative will arise at sites with high volumes) [while allowing variation as a function of traffic volume, for most rate groups the ADT adjustment factor is set to '0']

- Has a functional relationship that does not lend itself to modeling non-linearity
- Does not include variables other than volume as predictors of expected risk
- The one parameter that can be adjusted, the ADT factor, has apparently not been adjusted recently
- Implicitly assumes that all the factors causing high collision rates reside within the segment

Recommendation:

This method has some of the characteristics of other more advanced methods (see below), but is limited in its functional form. In addition, it appears that the one parameter that can be adjusted—the ADT factor—has not in fact been adjusted recently. We recommend that NE be replaced by more sophisticated methods (see below).

6.1.2. SAFETY PERFORMANCE FUNCTIONS

Safety Performance Functions are a predictive tool to estimate the safety of a highway site with specific design characteristics and traffic volumes. A safety performance function can be defined by:

$$N_e = f(AADT, x)$$

Where:

- N_e is the expected annual accident frequency
- AADT is the Annual Average Daily Traffic
- x are design characteristics and other variables

The procedure for obtaining a SPF has been described in detail in a SafetyAnalyst white paper (xx) and elsewhere (xx). The procedure involves identifying the appropriate functional form, identifying the significant variables, and calculating the parameters of the model empirically using data from a combined set of sites. SPFs have been successfully used in a wide range of situations (xx).

Note that the Table C formula has some similarities to the Safety Performance Function (SPF) in that it provides a relationship between frequency and volume. However, three differences are: **(1)** Table C assumes a fixed rate, although in some cases modified by traffic volume, **(2)** Table C has a different way of handling traffic volume (in Table C traffic volume is used as a factor adjusting rate, in SPF traffic volume is predictive variable in itself), and **(3)** Table C has no provision for including factors other than rate and traffic volume.

As an exercise we have constructed and tested a SPF for intersection data in TASAS. For this exercise we chose 3-legged intersections (xx). We produced several SPF models and compared them with Table C. In order to evaluate the model, we generated SPFs using data from years 1996–1999 and compared this to the to actual collision rates during years 2000–2003. The objective was to calculate the difference between the number of collisions predicted and the number of collisions observed during the years 2000–2003. Each of several different SPFs were superior to Table C predictions. In general, the superiority of the SPF over the Table C prediction was related to the greater amount of information taken into account by the SPF; the superiority of the SPF increased with its complexity.

Use of SPFs has become nearly a norm in determining HCCLs. One example of the application of SPH is provided by Kononov and Allery from Colorado DOT (xx). They have proposed using confidence intervals around SPFs to define a “Level of Service” (LOS) of safety for roadway segments. Collision frequencies beyond a particular confidence region would be considered high risk locations for further investigation.

The SFP provides several advantages over N_E in that it is a far more effective model for calculating the expected accident frequency and the associated distribution. The issue of serial correlation still persists for highway segments. The serial correlation arises because hot spots will tend to generate secondary collisions in neighboring sites. The resulting SPF could be shifted in one direction or another increasing false positives or false negatives. The magnitude of this bias is not known.

IN SUMMARY, THE SPF APPROACH POSSESSES THE FOLLOWING CHARACTERISTICS:

Strengths:

- Allows variation as a function of traffic volume
- Allows great flexibility in determining the relationship between number of collisions and traffic volume
- Allows inclusion of other variables defining individual sites

Weaknesses:

- Does not take into account actual collision counts at the individual sites in its modeling, as compared to the EB approach (see below)
- Not suitable for analyzing sites (such as urban freeways) where the collision numbers are not independent.

Recommendation:

- The method using Safety Performance Functions (SPF) should be systematically compared to Table C and Empirical Bayes where the collision counts are not correlated. (such as intersection and ramps) (
- The impact of serial correlation among sites should be evaluated.
- The impact of missing parameters should be evaluated.

6.1.3. EMPIRICAL BAYES ESTIMATE

The Empirical Bayes (EB) method is a method that combines two different types of information: the expected accident frequency based on experience in the entire set of comparable sites and observed frequency of accidents at a specific site. The expected accident frequency can be obtained by using the SPF calculated for the highway site. The observed accident frequency can be based on one or multiple years. The basic idea of the EB method is that there is important information contained in the actual observation made a particular site that is not used in generating the SPF.

Making use of two assumptions (such as accident frequency at a given site follows a Poisson distribution and the average accident frequency of comparable sites follows a Gamma distribution) a simple estimate of the site safety can be obtained using the Empirical Bayes method:

$$N = w N_E + (1-w) N_O$$

Where:

- N_E is the annual average expected accident frequency
- N_O is the annual average observed accident frequency
- w is the specific weighting factor to apply
- N is the Empirical Bayes Estimate

The weighting factor can be interpreted as a “trust factor,” as it indicates which of the two clues seems to be the most relevant. The weight factor is a function of the analysis period length, the estimated accuracy of the SPF, and the expected accident frequency. The formula of the weight is:

$$W = \frac{1}{1 + T \times k \times N_E}$$

Where:

N_E is the annual average expected accident frequency

T is the analysis period length

k is a characteristic parameter of the SPF (dispersion parameter)

It can be noticed that the weight decreases with the analysis period. Indeed, as noted elsewhere in this report, provided that the real risk of a site remains constant over the years, the longer the analysis period, the better the annual average approximates the real average. Consequently, if the analysis period is long, the weight is small and the Empirical Bayes estimate mostly uses the observed average accident frequency.

The Empirical Bayes Method can be applied very easily provided Safety Performance Functions have already been calculated. The procedure to obtain an Empirical Bayes estimate for a specific highway site is as follows:

- 1 Calculate annual average accident frequency over the analysis period considered using SPF.
- 2 Calculate weight using the characteristic parameter of the SPF used in previous step.
- 3 Calculate Empirical Bayes estimate using observed and expected average accident frequency.

The EB method has been used in a large number of applications. The important feature of the EB method is that, by combining the expected frequency generated by the SPF with the observed frequency, the regression to the mean phenomenon is mitigated. In some cases the gain over SPF is small (xx), but, given that it is fairly easy to calculate once a SPF has been established, it should be considered as a potential method for determining HCCLs.

IN SUMMARY, THE EB APPROACH POSSESSES THE FOLLOWING CHARACTERISTICS:

Strengths:

- Allows variation as a function of traffic volume
- Allows great flexibility in determining the relationship between number of collisions and traffic volume
- Allows inclusion of other variables defining individual sites
- Accounts for regression to the mean

Weaknesses:

- More difficult to calculate
- Less intuitive
- Not suitable for analyzing sites (i.e., urban freeways) where the collision numbers are not independent.

Recommendation:

- The method using Safety Performance Functions (SPF) should be systematically compared to Table C using TASAS data over a wide range of categories of sites.

6.1.4. THE ROLE OF PREDICTIVE VARIABLES IN SPFS (AND EB)

As described above, the SPF (and EB) model is very flexible in that a number of variables can be included in the model. The most important variable (most powerful predictor) is usually traffic volume, i.e., more vehicles usually means more collisions. There are three potential uses of this capability.

1 STANDARDIZATION (“COMPARING APPLES TO APPLES”)

Including traffic volume inherently “standardizes” for volume. That is, sites will be evaluated in relation to other sites with the same volume. Other variables entered into the model have a similar function. For example, adding a variable for shoulder width will in effect “standardize” for shoulder width. If shoulder width is inversely related to collisions, then the expected frequency for segments with low shoulder width will be “adjusted” upward. The general principal is the intent to compare sites with similar sites. When this is the intent, then the actual number of collisions at a site can be compared with that predicted by the SPF when all the variables have been set to that characterizing the site. This means that excess collisions (any amount by which the actual is greater than the predicted) are due either to noise (i.e., chance) or to some feature that is not available or at least is not used in the model. In fact, the role of further investigation would be to identify these features not included in the model.

There is one important implication. Some sites will be chosen with fewer collisions than other sites not chosen. Using traffic volume as an example, some sites chosen will have fewer collisions (with lower volume) than some sites not chosen (with higher volume). This is consistent with an assumption that the potential of reducing collisions is proportional to the excess collisions, and not to the absolute number of collisions. The actual result of this in terms of maximizing cost-benefit has not been determined (xx).

2 IDENTIFYING IMPACTS OF DESIGN

There is a danger that variables included in the SPF model might be neglected in terms of selecting countermeasures. For example, a model including shoulder width will permit comparing sites while controlling for shoulder width, identifying factors at each level of shoulder width that contribute to collisions, but taking the focus off shoulder width. However, the fact that low shoulder width is predictive of collisions means that low shoulder width is a design feature that should be addressed across the entire set of sites.

3 IDENTIFYING ROADWAY CATEGORIES

Another potential role for variables in a SPF is to assist in identifying roadway categories. The Table C method, and most other methods, begins by categorizing the roadway system into categories of similar types. The question can be raised, what defines “similar?” The SPF can be calculated using data within a category or within a cluster of categories combined. In former case, variables are identified to “standardize” comparisons among sites. However, in the later case, variables in the model provide a possible tool for defining categories. This can be done by calculating a SPF for two rate groups combined and then introducing interaction terms to determine if factors like traffic volume operated in the same way across the two rate groups. If so, then there would a rationale for combining the groups, and therefore increasing the size. An analysis demonstrating the feasibility of this approach has been conducted by combining different rate groups defining 3-legged intersections. This showed that it was possible to combine rate groups, resulting in a single rate group with a larger size and therefore leading to increased stability of expected collisions. It is suggested that this strategy be utilized in helping combine rate groups into large entities in cases where the numbers of sites are very small.

6.1.5. CONTINUOUS RISK PROFILE (CRP) METHOD

Continuous risk profile (CRP) is a new method for assessing collision risk along a roadway that addresses the limitation of a method that requires arbitrary segmentation of a roadway for analysis. Continuous risk refers to the concept that the road under examination is not segmented, but rather is considered as a whole. The method produces a continuous profile that shape of which reflects the true underlying risk along the roadway. CRP method has been developed by Chung and Ragland (xx) to be used by Caltrans traffic engineers. However, the general methods used for continuous risk profiling are applicable for any jurisdiction that examines collision concentration in urban freeway areas.

A CRP is developed in four steps: (1) calculating a cumulative count of collisions along the roadway; (2) estimating the excess risk compared to the reference risk defined by the user; (3) pre-filtering frequencies with small domain (i.e., the noise); and (4) profiling excess risk continuously along the roadway.

IN SUMMARY, THE CRP APPROACH POSSESSES THE FOLLOWING CHARACTERISTICS:

Strengths:

- Intuitive interpretation
- Does not require any changes in current Caltrans collision database.
- Does not require arbitrary segmentation of a roadway, but shows how risk varies continuously within or across segments
- Can identify secondary collision clusters (i.e., clusters of collisions arising because of congestion caused by collisions in a primary cluster)
- When estimating the effect of counter measures along the roadway, CRP captures the secondary benefit in the vicinity (i.e., reduction in collision rates in the adjacent sites) in graphical form.

Weaknesses:

- Not suitable for comparing collision rates at a short segment or isolated intersections.

Recommendation:

- The CPR should be systematically compared to other methods (Table C, SPH, and EB) where collision on highway segments (i.e., where collision counts are likely to be correlated).
- The impact of serial correlation among sites should be evaluated
- The impact of missing parameters should be evaluated.

6.1.6. COMPARISON OF METHODS

We have provided an account of strengths and weaknesses of four different methods.

Four methods for calculated expected frequency of collisions were compared. Table 2 summarizes the strengths and weakness of each.

It is fairly clear that the current method using (NE) should be replaced by more sophisticated methods, that some version of SPF or EB should be developed for intersections and ramps, and that there are two competing or possibly complementary methods for dealing with roadway segments.

Table 2

	Ease of use/ understanding	Allow for effects of traffic volume	Appropriate model
Table C (N _E)	Medium	Yes	No
Safety Performance Function (SPF)	Low	Yes	Yes for intersections and ramps Questionable for roadway segments
Empirical Bayes (EB)	Low	Yes	Yes for intersections and ramps Questionable for roadway segments
Continuous Risk Profile (CPR)	Medium	Yes	Yes for roadway segment

However, a number of questions remain:

- What form should the SPF take?
- How much is to be gained by developing an EB approach?
- What approach should be used for highway segments (SPF or CRP)?
- How will these new approaches be integrated into the current Table C system?

To answer these questions we recommend a pilot study to evaluate these methods side-by-side on a similar set of roadways. The proposal is as follows:

SITE: All intersections and freeway segments in D4

METHOD: Implement Table C, SPF, and EB at all intersections
Implement Table C, SPF, EB, and CRP at all freeway segments

PERFORMANCE

MEASURE: Ability to predict collisions from a set of base years to a set of target years

TIME FOR STUDY: One year

7. FORMAT AND CONTENT FOR REPORTING SITES

This section discusses the information to be included in the reports of HCCLs. The current Table C provides this list of information:

- Location
- Rate groups
- Total number of collisions different time intervals
- ADT
- Numbers of fatal and injury collisions

Even though it is desirable to have concise and brief forms of reports to be distributed to the users, there are advantages in enriching the outputs of HCCL screening for the benefits of assisting the users of Table C with additional and supplementary information.

Since the original database (TASAS) contains a much larger set of variables they can be used to provide helpful inputs for the follow-up evaluation and investigation. For example, by dissecting the crash records and performing post-screening analyses, the patterns, collision factors, and time history of crashes at identified sites can be compared to other similar sites. Furthermore, it will be ideal to link Table C to other existing database or data systems so that an integrated data system can improve the ease of use and overall efficiency. For example, if the results of Table C can be utilized in conjunction with a map-based Geographical Information Systems (GIS) then the distribution of collisions along a highway or in a region can be clearly visualized. For another example, if the follow-up actions of safety investigation and safety improvements can be linked to and tracked within archived or existing Table C records by inquires, it will greatly enhance the functionality of such reports.

The major observations and recommendations for these issues are provided in the sub-sections below.

7.1. INFORMATION PROVIDED

(E.G., HIGHWAY FACTORS, NON-HIGHWAY FACTORS, COLLISIONS FACTORS)

TABLE C:

Location, Rate Group, total collisions in different time intervals, ADT, and number of fatal and injury collisions.

Observations:

- Some states (e.g., Colorado) provide a much richer set of information about HCCL sites.
- Much more information is available in the TASAS that could be provided in the Table C report.

Recommendations:

- Expand the Table C report to include:
 - Information already provided
 - Collision patterns
 - Comparison of collision patterns to other similar sites (e.g., within the same Rate Group)
 - Provide trends over time at the site compared to overall trends at similar sites.
 - Other information that could be derived from TASAS or could otherwise be linked to the type of site and collision pattern

7.2. INTEGRATED DATA SYSTEM

TABLE C:

The Table C report provides fairly limited data in a list format.

Table C appears to be distributed as a somewhat isolated report, i.e., apparently with no systematic link to follow actions.

Observations:

- Providing Table C reports within the context of a broader data system may facilitate use and provide tracking capability.

Recommendations:

- Develop an integrated data system within which the Table C report is generated.
- The integrated data system would include:
 - Maps of Table C locations
 - Information on collision patterns available by pointing and clicking on a site
 - Tracking information including (i) results of investigation, (ii) installation of countermeasures, (iii) evaluation [i.e., pre-post collision history]

8. DATA QUALITY

Table C makes use of the TASAS (Traffic Accident Surveillance and Analysis System) database, which provides information about the California State Highway network. Variables in this system are important for identifying HCCL. The variables area are described in Appendix X (Transportation System Network [TSN], TSAR Reference Card).

There are three primary types of data:

- 1 Highway Inventory
- 2 Volume Data
- 3 Collision Data

It is clear that the quality and completeness of these various types of data is crucial to HCCL analysis. In general, we have identified several types of issues with the data. The implications and recommendations of these issues are discussed as follow.

8.1. IMPLICATIONS ON HIGHWAY INVENTORY

The State Highway System (SHS) includes more than 15,000 miles of highways, 14,000 ramps and 18,000 intersections. Variables include characteristics of the different types of sites.

There are four types of Highway Inventory variables:

- Standard fields (functional classification, highway group, etc.)
- Highway fields (lanes and other design features)
- Intersection fields (configuration, traffic control device, etc.)
- Ramps fields (configuration)

PROBLEMS:

There are four issues with the highway inventory data:

- 1 Missing design information for a small number of sites

Some variables have incomplete information but this problem is present for only less than 1% of the total data. No recommendation is made at this time. However, whenever such segments or sites are recognized in data processing and the relevant information become available, corrections should be made to enhance data sets.

- 2 A relatively small number of sties for some rates group

Some of the rate groups have a very small number of sites (see section XX). The implication is base rates calculated for these sites are likely to be very unstable. Rate Groups with small numbers should be combined with other groups.

- 3 Overlapping Sites

A small number of intersections and within 250ft of one another and collisions in between may be double counted. Intersections: double counting of accidents is due to the overlapping of the 'N' Area of distinct intersections. This overlapping of intersections' 'N' Area can cause problems for both calculating the expected accident frequency and estimating safety.

There are at least several potential approaches to tackle this problem:

- One approach is to identify the upstream or downstream direction of the roadway and associate the collisions to the upstream or downstream intersection only, when it is recognized that a second intersection is within a specified distance. This should eliminate the double counting problem.
- The other potential method involves the re-categorization of site types and an overhaul of rate groups. For example, if intersections are treated as a “segment” of a continuous roadway, then the calculation of safety performance will follow the use of the chosen methods in screening and identifying HCCL on a continuous highway.

4 Double Listing

A small number of highway segments and ramps are listed twice. These errors are minimal and should not affect the results. This is not a major issue and it will not affect the results. However, whenever such segments or sites are recognized in data processing, corrections should be made to avoid repetition of the errors.

8.2. VOLUME DATA

Traffic volume data are obtained from the Traffic Data Office (in Traffic Operations). Average Annual Daily Traffic (AADT) is available for all intersections, ramps, and roadway segments. The calculation of Annual Average Daily Traffic (AADT) is performed once each year based on data collected from October 1 through September 30. Volume is collected at all sites on a rotating basis once every three years. Using these traffic volume data, base rates for different roadway types are calculated in the following way:

- Highway Segments: Collisions/million vehicle miles
- Intersections: Collisions/million vehicles entering the intersection (primary + secondary)
- Ramps: Collisions/million vehicles traversing the ramp

8.2.1. PROBLEMS:

The research team has identified five issues pertaining to Volume Data:

- 1 Data often out of date, many data points interpolated, etc.
- 2 Some missing or out-of-range values

We found missing volumes for about 1% of highway segments, 1% of intersections, and 2% of ramps. In itself, this number of missing probably has minimal impact on Table C analyses but rates with this .values should be eliminated in any analysis.

- 3 Out of range volumes

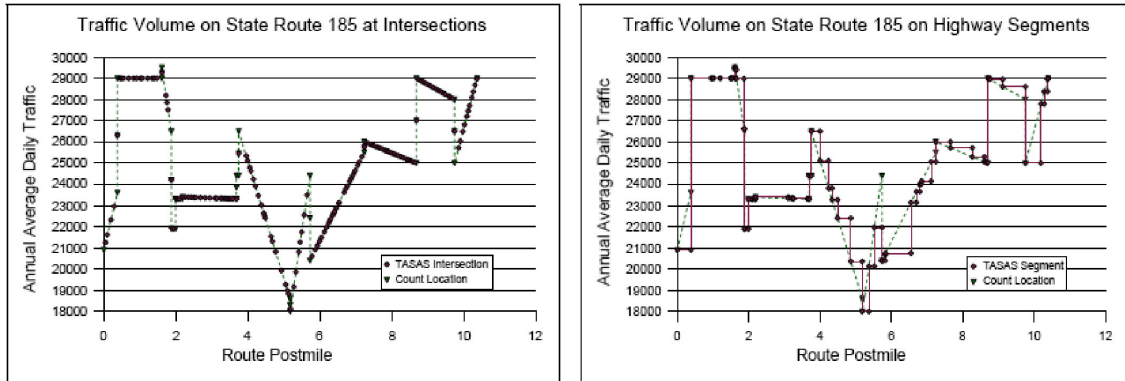
For intersections, we found that a fairly large number of intersections (about 5%) had very low AADTs (less than 10). A very small number of ramps (less than 1%) had very low AADTs). Such low volumes will result in very high estimates of rate, and could bias outcomes.

- 4 Interpolation or Extrapolation of volume estimates

The uncertainty in traffic volume information arises from the frequency of traffic counts used for estimating the AADT. Traffic volumes on state routes are recorded by Caltrans. In general, for each route, traffic counts at fixed control stations are collected once every three years. Based on a one (or several) day count and different factors, the Annual Average Daily Traffic is computed. Each year, the AADT is given for every control station whether it has been

Figure 4

EXAMPLE OF EXTRAPOLATION FROM RECORDED TRAFFIC COUNTS TO TASAS TRAFFIC INFORMATION

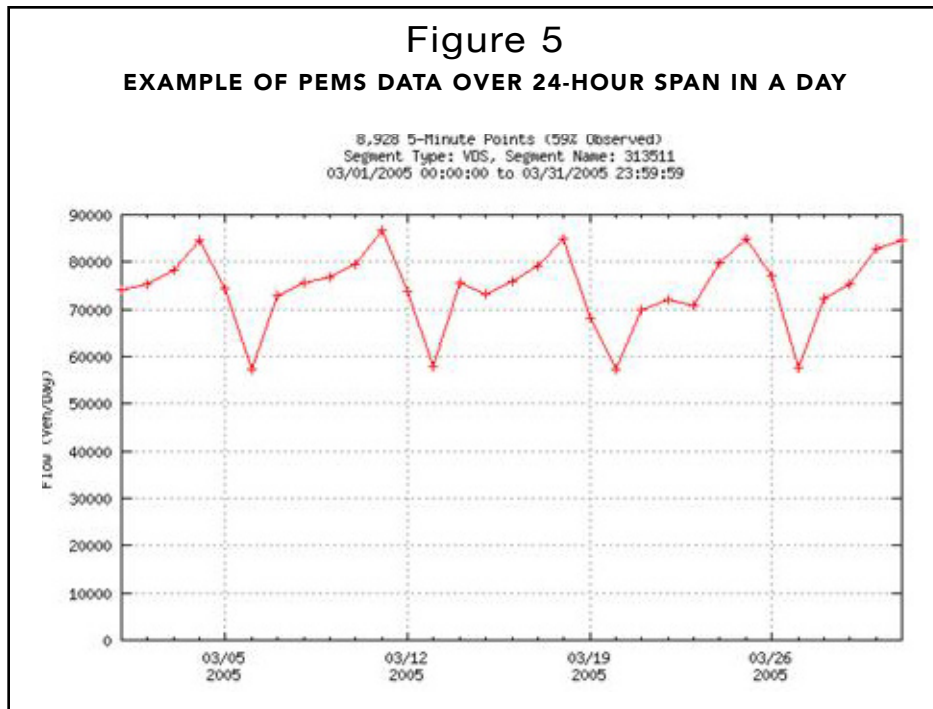


updated or not. The resulting tables are accessible online. They can be found on Caltrans website (8). Based on a few control station's AADT, the traffic volume for each segment, intersection and ramp is calculated in TASAS using linear interpolation. For intersection crossing roads, traffic volumes are obtained either by counts, using the same method as for State Routes but at a lower frequency (often once every 10 years), or by estimations. The extrapolation from the estimated traffic volume at few count locations to the traffic volume information coded in the Highway database is illustrated on Figure XX. For intersection crossing roads, traffic volumes are obtained either by counts, using the same method as for State Routes but at a lower frequency (often once every 10 years), or by estimations. Estimations are identified by a 1 for the last digit of the crossing street AADT and account for ~60% of attributed values.

5 Variability of bias in volume estimates

Currently procedure does not consider the effect of variations in traffic depending on different days of week and traffic demand. Suppose there are two sites with the same AADT where one site has high peak demand (typically observed in Northern California and the other with moderate demand that last through the day (typically observed in Southern California). The effect of the variation in traffic demand can have different effect on collision rates.

Figure xx shows the variations in traffic demand across different days observed on eastbound Highway-80 near the city of Roseville. The figure illustrates the fluctuations in daily traffic volume during a one-month period. The peaks on this chart occurred repeatedly on Fridays, when the traffic traveling in the Lake Tahoe and Reno direction



was considerably higher than the other days. The initial steps to take for the analysis of commuting related incidents will be to examine the number of incidents during selective hours of the day or selective days in a week. The total numbers of accidents or the distributions of accident types in the selective windows versus the overall distribution will provide the basis for evaluating the contribution of traffic volume and congestion related factors on the occurrence of incidents.

8.2.2. RECOMMENDATIONS:

Check TASAS database based on some of the results given previously:

- Add missing sites if appropriate.
- Screen sites with no accidents over a long period of time for closed roads or non-State managed roads (additional statistical criteria may be used to reduce number of sites to check)
- Check traffic volume information for sites with missing, incorrect or out of range values.
- Create methodology for checking TASAS (tests to perform and criteria).
- Feedback loop from Table C to TASAS to reduce number of errors
- Improve quality of traffic information data and reduce underreporting rate value and variance. For traffic volumes, it would be beneficial to consider two traffic volume fields, begin_adt and end_adt, if Table C can be made compatible with this update.
- Set up ongoing system to monitor quality of volume data and make improvements
- Develop statistical model of volume data to facilitate projects, interpolations, etc.

8.3. COLLISION DATA

Collision data are obtained from the California Highway Patrol (CHP) from a database called SWITRS (Statewide Integrated Traffic Records System) SWITRS is intended to include all police-report traffic collisions in the state. Collision data are extracted by CHP from the SWITRS database and contain information about collision aspects and party involved, coded by CHP, as well as site location, coded by Caltrans. Between 1994 and 2004, more than 1,800,000 accidents were recorded on Californian State Highways.

8.3.1. PROBLEMS:

There are four issues with the collision data.

1 Underreporting

Underreporting of accidents, that occurs when a portion of accidents are not reported, cause an underestimation bias in the observed accident frequency. Vogt and Bared (4) noted that “the amount of any underreporting is a matter of speculation (one source in Minnesota thought there might be one minor unreported accident for each reported one because accident-prone drivers wish to avoid both penalties for intoxication and insurance premium increases)”. A major concern is then to estimate the underreporting rate (number of observed accident divided by the real number of accidents). It is both important to know what is the underreporting rate and how it varies from an area to another. Indeed, if in certain areas the underreporting rate of accidents is smaller than in the other areas, then the corresponding highways will incorrectly appear safer.

2 Inaccurate information

The degree of inaccuracy is not known with level of certainty. Moreover, in analysis of location and movement preceding collision, we have found internally inconsistent information.

3 Linkage Issues

A small number of collisions could not be linked to a highway location (<1%). We have noted the following types of errors:

- Location Errors
- Errors in movement preceding the collision and direction.

4 Missed identification and underestimation Issues

One problem occurs when a segment in a Highway Rate Group that is less than 0.2 miles is currently ignored or not documented in the Table C and Wet Table C Overview. For example, if a Highway Rate Group is 0.5 miles long. If the first and second 0.2 miles segments are significant, then the last segment in the analysis for this Highway Rate Group will include 0.1 mile of the next Highway Rate Group. In this case, the analysis will stop and restart at the beginning segment of the next Highway Rate Group, and the last 0.1 mile of the previous Highway Rate Group will be ignored.

Another problem during Highway analysis appears when moving window is reaching the "N" area of an intersection—250 feet beyond the intersection. The analysis process will stop and restart beyond the "N" area, since accidents at intersections have already been analyzed in Intersection Analysis and will not be analyzed in the Highway Analysis. The collisions coded outside the intersection but within the 'N Area' (usually 250 feet) will have a File Type = 'H' however they are also included with the Intersection analysis. It means that some collisions are included twice as in highway file as in intersection file.

Due to the problems mentioned above, the implications for screening for HCCL are

- Some sites automatically considered as non dangerous by Table C
- In some cases, underestimation of expected accident frequency may occur.

5 Other Miscellaneous Issues

- In the accident file, some accidents are identified as "ramp" incidents, but their post miles fields are marked at locations before the post mile in the ramp file starts.
- In the accident file, there are ramp accidents that do not match any post mile in the ramp file.
- The highway accidents at some post miles fall in two segments of the highway data due to overlapping highway segments.
- There are intersection accidents that do not match with any location in the intersection data.

8.3.2. RECOMMENDATIONS:

- Create programs to do systematic range and missing value checks
- Prepare reports on out of range and missing data as feedback to CHP and other police agencies.
- Test models of extrapolation and interpolation

Some of the problems in collision data are associated with the reporting procedure, such as the under-reporting or missed information in the collision report. This is difficult to overcome due to the nature of the process involving human operators. However, other site specific errors if discovered in data processing should be corrected to avoid repetition of the errors in the future.

9. APPROACHES OTHER THAN SITE-SPECIFIC APPROACHES

9.1. INDIVIDUAL SITES VS. TYPES OF SITES

TABLE C:

Table C currently is designed to identify specific sites, such as intersections, ramps, and 0.2 mile segments.

Observations:

Methods such as Table C focus on comparing sites with common characteristics to identify those which have a high number of collisions in relation to other similar sites. However, when such sites are identified, they would be necessary have some characteristic(s) that differentiates them from the other sites that generates the high number of collisions. Such a characteristic(s) is often a design characteristic that may in fact appear in other sites also with a high number of collisions. This suggests the strategy of identifying not just specific sites with high risk, but design features with high risk. The methodologies of SPFs, EB methods, and the CRP method all lend themselves to implementation of this strategy. Parameters in SPFs can represent design characteristics (e.g., shoulder width, curvature) that affect collision risk, and that could be addressed on a large scale (i.e., not just as a feature of a specific high risk site).

Recommendations:

Statistical models such as Safety Performance Functions, Empirical Bayes Methods, and the Continuous Risk Profile method, should be developed to identify patterns of collisions related to various design features.

9.2. CORRIDORS

TABLE C:

Focus on individual sites

ObservationsOne of the findings reported in the Table C Task Force Report is many required or recommended highway segment locations were in fact adjacent. One of the recommendations was to combined adjacent locations which would create segments up to 1 mile. In fact, we have found that various methods of identifying high collision sites will often yield adjacent locations. The phenomenon is not limited to highway segments. In many cases, neighboring intersections may have concentrations of collisions.

With the current Table C method, clusters are adjacent sites are all based on noted patterns among sites selected because of high risk collisions in themselves. However, (i) there are some reasons to believe that traffic collisions may be affected by common factors within a large area that a single intersection, ramp, or 0.2 mile highway segment (ii) areas may have common features (e.g., such as non-optimal signal timing) in a number of related sites, and (iii) some countermeasures may be more effectively implemented across a set of sites or within a community. In other words, in some cases the most appropriate "unit of analysis" may be broader than a specific site.

There are in fact methods for identifying sites larger than 0.2 or for identifying clusters of specific sites (such as intersections). Several approaches include: (i) using a "sliding window" of different lengths; (ii) utilize a method similar to Table C but choosing a much larger interval (e.g., 1/2 mile), (iii) calculating collision frequencies (or rates) in highway segments larger than 0.2, and (iv) calculating continuous densities of collisions by plotting collisions using GIS methods and then using existing software to calculate clusters, or regions that show a high level of collision density.

Recommendations:

It is recommended that, as supplement to the Table C program for identifying specific sites, that Caltrans develop and implement a parallel methodology for identifying clusters or "corridors" with a high collision density and that this be part of the regular Table C reporting.

APPENDICES



Traffic Safety Center

Setting New Directions in Traffic Safety

