DRAFT TACT REPORT – METHODOLOGY MANUAL

A National Model for the Evaluation of CMV Selective Enforcement Programs

Methodology Manual

Dr. David Brown, Dr. Randy Smith, Dr. Brandon Dixon
Center for Advanced Public Safety (CAPS)
Dr. Steven Jones
University Transportation Center of Alabama (UTCA)

The University of Alabama

for
Alabama Department of Public Safety (DPS)

and
Federal Motor Carrier Safety Administration (FMCSA)

DRAFT METHODOLOGY MANUAL

September 30, 2011

TABLE OF CONTENTS

- 5.3. Gather and summarize administrative data after the project
6. Analytical techniques (statistical analyses)
    - 6.1. Crash data analysis
        - 6.1.1. Student’s t-test method (using Excel)
        - 6.1.2. Binomial probability method (using Excel)
    - 6.2. Citation data analysis
    - 6.3. Survey data analysis
    - 6.4. Observational data analysis
7. Remarks on drawing conclusions
    - 7.1. Accomplishing the evaluation objectives
    - 7.2. Proper use of qualifiers

1.0 Introduction and Background

This document (henceforth referenced as the *Methodology Manual*) is intended to provide a quick reference to the more detailed examples for this project entitled "A National Model for the Evaluation of CMV Selective Enforcement Programs – Supplemental Report." That report, which will henceforth be referenced as the *Supplemental Report*, contains extensive details and examples to define the procedures recommended in evaluating Commercial Motor Vehicle (CMV) selective enforcement programs, and in particular, the Ticketing Aggressive Cars and Trucks (TACT) program.

For brevity and convenience in this manual the acronym "TACT" will be used to apply to ***all selective enforcement programs involving CMVs*** even though they might not be designated (funded) as formal TACT programs. The intended audiences for this *Methodology Manual* are law enforcement supervisors and their support staffs (including consultant services) who share the duty of improving programs involving CMVs through the use of data and analytical methods applied to problem identification and evaluation. Also, it must emphasized that the findings of the recommended evaluations should lead to improvements throughout law enforcement selective enforcement and PI&E operations, and not just TACT programs, or for that matter, those involving CMVs. So, unless otherwise seen from the context, the acronym *TACT* is used in a very broad sense throughout this document and the Supplemental Report to refer to all similar programs involving CMVs.

TACT programs are primarily interested in reducing the conflicts between CMVs and other vehicle types. In this report the word "cars" is used to collectively represent all of these "other vehicle types." This is done for brevity recognizing that there are many non-CMV vehicle types that would qualify, including sedans, vans, mini-vans, etc. A crash between a CMV and a car (or a car and a CMV) is referred to as a CMV-car crash. In using this term (CMV-car), there is no implication as to which of the two vehicles (or their drivers) caused the crash. The term CMV-car is used throughout this document for consistency, but it is undistinguishable from car-CMV; the two should be considered interchangeable in the context of this manual. When fault is to be attributed to one of the drivers, it will be stated explicitly.

1.1 Purpose, Mission and Strategy

The purpose of this document is slightly different from that of the overall project, which was to develop the methodology for TACT evaluations and document it as comprehensively as possible. The primary goal of this document is to provide an easy point of reference to that project Supplemental Report. One of the primary objectives in creating the *Methodology Manual* was to briefly outline the methodology for practitioners interested in performing evaluations apart from the detailed examples. All users do not need all of the information in the Supplemental Report. Some are already quite knowledgeable on many of the aspects of problem identification and evaluation techniques. Thus, this document is an abridgement of the Supplemental Report intended to serve primarily as a checklist, and secondarily as a reference back to the Supplemental Report when more information is needed.

1.2 Discussion of Motivation and Optimization

Definitions. The word *program* is used here to refer to an overall traffic safety effort (e.g., the TACT program within a given state) which might be funded for a year. Within each program there are generally a number of *projects* that are conducted (sometimes called *waves*). The various *waves* of selective enforcement and Public Information and Education (PI&E) within a given TACT program would be considered to be projects.

Traffic safety has suffered immeasurably from evaluations that were motivated by very little other than a desire to prove a program to be worthy of continued funding. The methodology presented herein is based on the position that overall funding decisions should not be based solely on the evaluation of one project within a total traffic safety program. If at all possible, all projects need to be evaluated and their costs and benefits assessed so that the total program can be optimized. The primary motivation for evaluating any given component (e.g., a TACT project) must be that of *improving project effectiveness in the future.* "Improvement" in this context might consider major re-direction of future resources, but generally it will be geared toward minor modifications that will continue to enhance the projects in future implementations. If all traffic safety problem identification and evaluation could be viewed in this context, it would lead to greater objectivity and dramatically improved traffic safety programs. And, of course, what applies within a program would apply to the total statewide traffic safety efforts in determining optimal tradeoffs among programs.

There are even tradeoffs within a given TACT program, and the data obtained from the evaluation should be targeted to helping in attempting to obtain an optimal balance in allocating resources. Three obvious competing entities are (1) selective enforcement, (2) PI&E, and (3) the evaluation itself. The total TACT program budget has to be allocated to these three items and any increase in emphasis on the one will decrease the efforts on the other two. The balance between selective enforcement and PI&E is a difficult one to make, and one of the goals of evaluation should be to assess and optimal level of PI&E.

It is obvious that there can be an over-emphasis on evaluation, and the question as to how often projects should receive intensive evaluation is an important one. Most of the published TACT evaluation efforts have been fairly expensive, obviously drawing funds and other resources away from supplementary enforcement and PI&E efforts. The following considerations might assist in this decision making process:

- The value of evaluation is in the improved policies and tactical decision-making that will follow. While the value of these items are extremely difficult to assess, a general rule is that it is well worth paying from 5 to 10% of the total price of a program to improve the decisions made within that program.

- Consider the extreme: without some level of evaluation there cannot possibly be any sense of control in the various approaches that are being applied in guiding the details of the TACT projects.
- Consider the other extreme: if an intensive evaluation is done for every TACT project this will obviously consume extensive resources that would be much better employed directly in the project itself or elsewhere.
- The solution lies in the understanding that all evaluations are imperfect samples of the past. The extent to which these past samples are able to guide future decisions is critical to the assessment of how often to sample. If one good evaluation in the past has established the value of a countermeasure, then that would be adequate until some conditions change enough to make its conclusions invalid.
- The other factor in this decision is the cost of the evaluations. Some evaluation data are quite easy and inexpensive to obtain, and perhaps will be available and can be used for practically no cost (e.g., crash data). Other special studies that require extensive instrumentation might make frequent measurements prohibitive.

The above suggests a two-level approach to evaluation in general. Persistent measurements and analysis should be made when the cost is low and the information derived from these measurements is quite valuable in guiding the day to day operations of the program. More expensive special studies are warranted, but on a sampling basis and with much more planning.

Given this motivation, the following are *objectives* that might be set to be accomplished by the TACT evaluation process:

- To determine the benefits of the specific project and to establish the best estimate of its effectiveness in terms of reduced crash frequency and severity as well as positive changes in driving behavior;
- To attempt to find at least one weakness in each of the project components;
- To overcome these weaknesses by formulating recommendations for future projects; and
- To seek out and establish, if possible, new and creative strategic approaches toward reducing the frequency and severity of CMV involved crashes.

The objectives do not include justification; a continuously improving program will use evaluation techniques to seek out weaknesses as well as to highlight strengths. Ultimately, it is hoped that successfully meeting these objectives will stimulate efforts to integrate the lessons learned from the evaluation into all of the law enforcement day to day operations.

1.3 Background Sections

This section of the Supplemental Report presents examples of the type of introductory information that might be included in the introductory section of a TACT evaluation. Generally the background material should cover enough of the history of the program to enable the evaluation to be placed in its proper context.

2.0 Guide to Using This Document

As indicated above, this document is intended to serve as a checklist for project administrators and evaluators. *The section numbers of this document are generally identical to the corresponding section numbers in the Supplemental Report,* thus providing a direct means for referencing additional information (e.g., instructions and examples) on any given subject in the Supplemental Report. Generally, no additional reminders will be offered to the reader that this additional help is available.

The topics are considered in the same (chronological) order as they are expected to be implemented for any particular *project.* It is expected that the entire TACT *program* that is implemented by a state will consists of several such projects, often referenced as *waves* or *details*. Programs (consisting of a series of projects) are usually defined by an allocation of funds to TACT for a given period of time. The approach toward evaluation given here applies to one project at a time, and not necessarily to the entire program (depending on how large or over what period of time the program is implemented). The examples given in the Supplemental Report contain projects from the largest possible (i.e., statewide) implementation to one that is extremely localized and set up specifically to evaluate a very small project.

The evaluation itself should be viewed as consisting of two separate entities:

- Administrative evaluation – a series of ongoing measurements to verify that the projects (and ultimately the program) accomplish the activities that were specified in the project plan (e.g., 500 hours of TACT-related overtime). For additional administrative purposes, data on the actual activities will be accumulated (e.g., names involved, times, locations, etc.), and this information could be extremely useful in the effectiveness evaluation. For example, the specific times, officers involved, and locations might be applied to determine the particular citations that were issued as a result of the program.
- Effectiveness evaluation – measurements to determine the effectiveness of the particular project (or combination of projects) in bringing about a specific project objective. For example, a project objective might be to reduce the proportion of CMV-related multiple vehicle crashes that are caused by young drivers.

For funding and accountability purposes, generally all projects within a program will be given the same administrative evaluation. However, program resources allocated to the effectiveness evaluation will generally not allow all of the projects to be given the same intensive consideration. Thus, considerable up-front planning must go into determining the particular evaluation metrics that will be applied and the set of projects to which they will be applied.

The sections below are generally in chronological order, starting from the planning process before the project(s) to be evaluated and continuing through to the analysis of the evaluation metrics. *The one exception to this rule is the current section* (Section 2.0), which includes a Literature Review in the Supplemental Report. Methods for conducting a literature review are beyond the scope of this project; suffice it to say that the Internet search engines make performing a literature review quite easy, and it is very important that those doing evaluations be familiar with the latest complementary works that are have been performed.

3.0 Preparation for the Project

This section covers those steps that must take place prior to the implementation of the TACT program to assure (1) that the TACT projects/program itself is well planned, and (2) that sufficient consideration is given to evaluation. This will start with considerations for the planning process itself, starting with the selective enforcement, the PI&E and the evaluations. This will be followed by problem identification steps that should precede the actual TACT selective enforcement effort. The actual pre-project data gathering efforts are given next. Finally, considerations are provided for the necessary data gathering for the administrative evaluation.

3.1 Establishing the Project Plan

The overall project plan can follow the overall checklist given below. It will be expected that some planning documentation will have been developed to qualify for funding. This basic planning document will be augmented to consider the most recent historical data that are available and their application to generate the best possible set of projects. The following subsections provide a checklist for the planning effort.

3.1.1 Plan for the Selective Enforcement Project Itself

This is a good place to start the planning effort as it is likely that documentation exists that includes a preliminary plan that was assembled as part of the proposal. It is also a good place to start because it is the center of effort. As detail is added to the plan, it must be recognize that there will be problem identification efforts and perhaps recent developments that may ultimately re-shape the plan. So while the development of an outline plan is presented in the following sections, it is not intended that all issues be resolved at this stage in the process. The plan needs to be a dynamic working document that will continue to be improved right up until implementation. Prior to developing the detailed plan the following should be considered:

- Who will be involved? Include the management and officers who will be conducting the selective enforcement details as well as those performing the evaluations.
- What specifically will they be instructed to do? In particular, how will the TACT project be different from their normal details?
- How much time will each of them be allocated to accomplish their tasks?

- Where will they be required to perform these activities? (This may be deferred until the problem identification steps described below.)
- When will they do it? This should be specified both in terms of a time range (of weeks or months) and also specific times of day (also part of the problem identification).
- Are there any special support resources that they will need (e.g., training)?
- What administrative requirements will need to be met? (See administrative evaluation recommendations below.)

The purpose at this point will be to establish and overall skeleton for the plan and then as information is obtained relevant to the various questions above, the "blanks" will be filled in.

3.1.2 Plan for the Public Information and Education (PI&E) Effort

The questions for the PI&E plan are essentially the same as those given above for the selective enforcement itself. However, since there are usually a different set of individuals assigned to accomplish the PI&E, it is beneficial to separate these in the plan. As many details as are known at the outset should be documented, but usually the plan cannot be completed without further information.

3.1.3 Plan for the Use of Crash, Citation and Survey Data

This step involves the initiation of planning for the evaluation efforts. As with the actual SE and PI&E efforts, this step can only be completed as information becomes available. Nevertheless, planning to the extent possible must be initiated well before the project starts since the general approaches to each of the evaluation components need to be given consideration that might affect the implementation. Plans can be subdivided by the following data types that will be necessary to evaluate TACT selective enforcement and PI&E efforts:

- Crash data and Citation data – virtually all of the considerations for crash data apply equally to citation data, and these are covered in Section 3.1.3.1.
- Survey data – This is covered in general in Section 3.1.3.2. The design of a survey form is not trivial, and additional considerations will be given in Section 3.6.
- Observational data – this is highly dependent on the type of observational data to be collected, and it requires considerable forethought; this is covered in detail in Section 3.1.4

The plans for the data collection and analysis need to be developed at the same time as the plan for the project itself. While the process is presented in chronological order, the thought process for defining the data requirements must be conducted in reverse of this. By this it is meant that to define the data to be gathered, it is imperative that the statistical analysis be defined as a prerequisite. Far too many traffic safety evaluation projects are not adequately "thought-through" in advance. If the evaluation is an afterthought it will lead to a thought process such as this:

> "Oh, now that the project is over we have to perform an evaluation. Let's see now, what data do we have that we can use? Given these data, what statistical analysis techniques can we apply?"

Planning is emphasized from the outset to avoid this quandary. Ideally, early planning will be conducted in the following sequence:

1. What ultimate metrics are needed to both assess the effectiveness of the project and to guide future improvements?
2. What statistical techniques can produce these metrics and assure that their significance is determined?
3. What statistical techniques would be most appropriate – including what types can be readily performed by the evaluator(s)?
4. What data are needed to drive these statistical techniques? The answer to these three questions enables the following question to be answered …
5. What data collection instruments or software need to be developed to provide these data?

Again, while the details will be covered below, consideration needs to be given to the above issues at this point. Thus, it is strongly recommend that this entire manual be reviewed, and especially that Sections 5.3 and 6 be given special consideration. Once the analytical techniques and data requirements are understood, then the plan for their collection can be effectively developed.

Since observational data is usually customized to the state's specific program and data collection capabilities, it is treated separately in Section 3.1.4.

3.1.3.1 Crash and Citation Data

In most states crash and citation data are fairly well fixed by the systems that are in place. Since crash data systems have been a requirement of NHTSA since the early 1970s, these systems are fairly mature. In many cases this is their downfall, and the latency in getting paper forms processed and into them is problematic. For example, if it takes six months to officially "close out" a year, then it might be impossible to perform any type of timely evaluation using crash data. By the time that the evaluation data are available and the analyses are completed, so many changes have occurred that the results might not be at all useful.

In states that have such delays special efforts might be required to short-cut the process in order to obtain the data necessary for the TACT evaluation. The FMCSA requires that special data be collected for those CMV that qualify according to severity. It is possible that these reports could be either used per se or extended to all CMV crashes in order to provide information for the TACT evaluation. If this is necessary then action needs to be taken well before the TACT selective enforcement and PI&E efforts in order to assure that the "before" data will be available.

By the time this report is published most states will either have or be considering an electronic crash reporting system, and many states already have an electronic citation system in use. The problem with both of these is that it is difficult to turn them over immediately so in some cases part of the data will come in on paper and the other part electronically for at least a few years. This can be further exacerbated if the electronic and paper data are in different formats or structures. Such cases require significant additional planning in the early stages of the evaluation.

3.1.3.2 Survey Data Considerations

Example (model) survey instruments are available in the Supplemental Report and also from several states' TACT web sites, e.g., Washington State (19). It is quite important that the surveys be as brief as possible and still accomplishes their goals. An overwhelmingly long survey will be skewed toward those who don't mind being inconvenienced by it – others will either not get into it or quit halfway through. If a respondent gets aggravated at the survey, his/her response will be to the survey and not to the project being evaluated.

The goal of the survey needs to be clearly defined with regard to each respondent. For example, the primary goal of a survey to police officers may by to solicit their suggestions for program improvement. Dissatisfaction about a particular aspect of the program may well surface the fact that officers feel like their time could be spent more effectively taking a different approach. On the other hand, a survey of the general public might have the goal of determining if a particular effort is effective in communicating a favorable message to the general public.

The raw results of a single round of surveys are not nearly as informative as before-after or test-control comparisons. The relative change or difference in a given question response can indicate whether the program is having a significant perceived effect. Irrelevant questions should be avoided in the interest of brevity. On the other hand, considerable information can be obtained from demographic data (e.g., age, gender, race, driver type, miles driven per year). While they are not related to the TACT program, they can be used for creating subsets and performing cross-tabulations once the results are obtained. Recognize that if such sub-stratifications are planned, then the number of survey forms will need to be increased appropriately. A rule of thumb is to have subsets no smaller than 30, although practical considerations may prevent this in all cases.

Driver type should include an indication of whether the driver is, or has ever been, a truck driver or a police officer. This is not to say they should be excluded from the general population surveys; in fact, it could be interesting to compare their responses to those obtained from the general population.

Since the target groups might impact the questions to be included on the surveys, they should be defined first. Three possible target groups are suggested: (1) law enforcement officers who have an understanding of the TACT program, (2) truckers, and (3) drivers in general. Generally the officers and truckers will be quite knowledgeable of the TACT program without either the enforcement or the PI&E, so the surveys administered to them will generally be of a different nature than that given to drivers in general. Surveys to the general public might best be restricted to drivers, since their attitudes must be impacted if the TACT program is to have any effect.

Drivers in general may be further subdivided into target subgroups. For example, in the younger driver initiative it might be required to generate additional submissions from those of the age group under consideration. It might be difficult to administer a survey to a particular target group such as this and considerable creativity is recommended in taking advantage of available resources, such as advocacy groups or college campus contacts, in reaching the target group. Alternatively, a large enough group might be surveyed at random to enable the subset with the target attributes to produce the statistical significance required.

In summary, the following survey types are recommended (numbers given below indicate the section in which examples are given in the Supplemental Report):

- Driver survey (3.1.3.2)
- Law enforcement survey (3.1.3.3)
- Trucker survey (3.1.3.4).

3.1.4 Plans for Conducting Observational Studies

Observational studies require considerable planning effort if they are going to be objective and return effective results. Raw numbers alone are rarely comparable before, during, or after the TACT effort. There are too many variables, such as traffic flow and even the number of CMVs on the roadway at any given time. This can be solved by transforming the counts into a rate by dividing by a count of potential occurrences (e.g., the number of tailgating events per hundred vehicles observed). So, along with a clear definition of what constitutes a hazardous incident, there must also be a definition of what constitutes a "non-hazardous occurrence." By recording both types of events it will be possible to come up with a rate = unsafe events/(unsafe events + safe events). These types of considerations will be covered in the following sections.

3.1.4.1 North Carolina TACT Evaluation – see Supplemental Report

3.1.4.2 Approach of the Alabama Example – see Supplemental Report

3.1.4.3 What Events Will Qualify?

The approach of the TACT program is to reduce certain types of hazardous events from occurring and thereby reduce the resulting crashes. In order for useful event data to be used for evaluation, the particular hazardous events must be defined as objectively as possible. For example, "following too closely" is highly subject to interpretation. Even following within a given number of feet may lead to considerable variation among data collectors. The following rules apply to defining events that qualify as "unsafe of hazardous events:"

- They must be specifically addressed by the program itself. If not, no variation can be expected. Clearly it is a hazard, for example, for a truck to linger in the passing lane. Question: is this in any way covered either in the PI&E or the selective enforcement? If not then it does not warrant measurement since no behavioral change of this type can be expected from the TACT project itself.
- Hazardous events must be easily and clearly identified. For example, if truck blind spots are well defined then a vehicle staying in such a blind spot without relative movement for greater than a given time interval (e.g., 20 seconds) can be considered a hazardous events. An observer would not be expected to use a stopwatch. It should be clear, for example, whether a private vehicle is lingering unnecessarily in a blind spot or trying to move out of it.
- Non-hazardous events must also be defined. For example, all observations of CMVs may be placed in one of the three categories: hazardous, undetermined, and non-hazardous? The rate would be determined as “the number of hazardous events per observed CMVs” (or perhaps per 100 observed CMVs). It would be perfectly acceptable, however, to exclude those CMVs from the observations when there is no car in their vicinity. If this were done then the rate would be “hazardous events per observed CMV that is in the proximity of a car.” Of course, considerable problems will arise if all observers are not recording the events and occurrences consistently and according to the definitions that are given during the planning stage.
- Allowance must be made for traffic conditions, and consistently moderate conditions should be sought for all data collection. As an extreme example, when traffic is at a stand-still, it is not a hazardous event for a car to be in a truck’s blind spot for any length of time.
- Ambiguous events must be handled consistently. One alternative would be to discard all questionable events. In other words, if an event is marginal and the recorder cannot make a clear decision as to whether it is hazardous or not, the event would not be recorded at all, and thus it would be added to neither the numerator nor the denominator of the rate formula. An alternative that might be superior would be to allow the raters to have a mid-entry when it is not definitive that an event is either safe or unsafe. This has the advantage of not losing any information on events, and then number of these “neutral” events might be added to the denominator. Perhaps an even larger advantage is in comparing the different raters for consistency. If the number of neutral ratings for any given rater is inordinately large, this would be an indication of indecisiveness, and perhaps the need for additional training.
- If possible both training and data collection should be aided by video recordings.

With these guidelines a plan can be formulated for the observational component of the evaluation. Recognize that consistency in measurement, and especially consistency among observers, is more important than a perfect definition of an unsafe event. Since the rates are going to be compared with each other (before-during-after), as long as the definitions are applied consistently, the metric should be effective in determining if the project is having a significant effect.

3.1.4.4 How Will The Events Be Recorded?

Data collection forms (preferably computerized) should be set up to assist the observers in obtaining uniform data. The form should contain a tally of hazardous, indeterminate, and non-hazardous events for each of the various types of hazardous events that have been defined. In addition to clear definitions for the various hazardous events to be recorded, the plan needs to contain the procedures for recording and forwarding the data collection forms. If at all possible the observers should be objective and impartial, e.g., ideally, there is no reason that they should know whether they are gathering data from a test or a control area.

3.1.5 Plans for Special Emphasis Area Analyses

The term *special emphasis area* refers to efforts targeting a particular classification of crash, such as those caused by young drivers. Other examples: specific roadway types (e.g., Interstates), those with high severity, those caused by speed or DUI, or any other target area. Most TACT projects do not have any special emphasis areas, and it that case this particular step in the planning process can be omitted. However, if there is some special emphasis area, then an analysis should be performed to determine any alteration in the standard TACT strategy that should be applied. Generally this can be accomplished by comparing the subset of crashes that includes the emphasis area against that subset that does not include the emphasis area. For example, compare CMV crashes caused by young drivers against CMV crashes caused by other than young drivers. Variables that show significant differences may surface times, locations, contributing circumstances, impact speeds, alcohol/drug use and many other factors that could help to target the TACT effort. At this point such should be worked into the plan so that this step is not omitted during the other problem identification steps.

3.1.6 Administrative Evaluation Plans

If there is already an adequate administrative data collection process in place, then plans will not be needed at this point to create one. However, if this is a new TACT project, or if the administrative data collected in the past was inadequate, then plans should be made and detailed to the extent possible at this point before going on. This might include the development of hard-copy forms or the preliminary design of an administrative data entry system. Administrative data has two goals: (1) to satisfy the reporting and accounting requirements of the TACT grant, and (2) to provide some key parameters of the effectiveness evaluations. The first of these is usually dictated by the grant funding requirements itself, e.g., number of details, number of citations issued of each type, etc. This is basically summary information (by month or quarter), and it may not include some key elements needed to support the effectiveness evaluation. This problem is further exacerbated by the delay that often occurs between project completion and evaluation necessitated; for example, the delay in getting crash data into the system. By the time that these data are available, data as to the exact start, end, and locations may be lost if not part of the formal administrative data gathering effort,

The following data elements should be considered in the administrative evaluation planning step to assure that sufficient information is available to support the effectiveness evaluation:

- The specific times and locations of the selective enforcement effort. The detail depends on the type of evaluation that is planned. For example, if the crash effect is to be measured at the exact times of the selective enforcement, then exact times to the hour and specific locations to the milepost would be required. More detail is always preferable to less as long as its collection cost is reasonable. In all cases the times should be coupled to the locations.
- The officers involved. These would also be coupled to the time and the locations for accounting purposes and to determine if there are significant differences between officers.
- Any particular emphasis of a given detail. For example, if a particular detail involved a ride along of an officer in the truck, this would be a different approach than the typical selective enforcement. Also, if the enforcement is going to be restricted to particular types of violations, that should be noted.

This concludes Section 3.1, which has been primarily involved with the planning process prior to a given project. The next section deals with applying existing data and analysis for the purpose of improving the effectiveness of the projects.

3.2 Perform Problem Identification

Problem identification is a term used in traffic safety with regard to investigating all aspects of a given subject area to determine just what the underlying causes are, their relative occurrences, and how they can best be addressed. The "subject area" of the TACT projects starts with CMVs. However depending on the projects' goals, it might get much more detailed in investigating particular aspects of the major area, e.g., youth-caused two-vehicle crashes on specific roadways.

Problem identification is an essential prerequisite to an effective program. There is no way to guide the tactics of a project without it. It is true that some experienced field officers often have as much if not more capabilities to guide a TACT program than can be attained by the data alone. However, this is no reason to ignore the potential additional information that can be attained from a formal problem identification step. It is quite difficult to merge all of this information and to document the results. Further, it is quite rare that a comprehensive look at the data in a formal problem identification process does not turn up some counterintuitive findings.

In the subsections below the problem identification will be subdivided into (1) hotspot analysis, which is mandatory for all TACT projects, and (2) supplementary problem identification, which would only apply to those projects that are further targeted on a particular crash type (e.g., those caused by 16-25 year old drivers).

3.2.1 Site Selection – Hotspot Analysis

Most states have a high-crash location capability. For example, they can find the location that has reported the most crashes over a given period of time. Or, alternatively, they can generate a list of locations that rise to the top in crash frequency and crash rate. In addition, most have a filtering capability whereby they can apply this capability to any subset of data, such as crashes involving CMVs. Locations so identified (according to whatever criteria of rate or frequency the state should choose to apply) are referred to as *hotspots*. It was stated above that such an analysis was mandatory for all TACT projects based on the fact that to have such a capability and not use it would border on negligence.

The following are recommendations with regard to using hotspot analyses in an attempt to maximize the effectiveness of the TACT program:

- Unless the data indicate otherwise, use at least three years of data in determining hotspots. CMV crashes are relatively rare, and multiple years are required to improve the determination of where clustering occurs on the roadways. As long as no major changes have occurred as far as CMV traffic is concerned, additional years over the three-year minimum may be added. However, at a certain point the adding of another year from the distant past will tend to be counterproductive. Remember, the goal is *not* to know the past – the goal is to predict what is going to happen in the future, all other things being equal. Studies have found that for predicting the location of high crash locations in subsequent years, three years is optimal. However, this study has not been repeated for CMV hotspots.
- Filter the data prior to running the hotspot analysis to just CMVs. It is interesting to see where CMV crashes deviate from all crashes. That could indicate some special problem either with CMV propensity to crash or it might just be a concentration of CMVs. The latter cause should not be discounted … clearly to counter CMV-involved crashes it will be necessary to go to where there are concentrations of CMVs.
- Subsequent filters for hotspots might include severity. Clearly the concentration of total crashes is going to be in or around the larger cities, but this is generally slower traffic and the crashes in urban areas are typically of lower severity. Using some combined injury and fatal crashes is probably as good if not better an indicator of where fatalities are likely to occur in the future. Basing decisions strictly on fatality locations is not advisable. Compare the high severity hotspots with the all-crash hotspots to get a feel for the problems with regard to injury and fatality crashes.
- The more filtering that is done the smaller the number of crashes at any one location, so the criteria for hotspots will have to be adjusted appropriately. When the number of qualifying crashes at a hotspot gets down below five, it is questionable if this is statistically valid, although the analysis may still be quite interesting and combining hotspots might still be useful in establishing patterns.

- Use your hotspot analyses not only for targeting selective enforcement but for determining optimal locations for PI&E countermeasures, such as the location of billboards or the placement of electronic signs.

3.2.2 Supplementary Problem Identification

Problem identification can be considered as a "fishing expedition" that involves searching through data to find out anything that might be useful for the issue at hand. It is recommended that, if possible, the agency run comparisons of their subsets of data for all variables in the database; or if that is not possible, then the key variables that are of interest be selected and used for the comparisons. The following comparisons will surface any over-representations that would be indicative of major differences that might be of use in customizing the TACT approach.

- CMV crashes against non-CMV crashes; this will surface variables at the highest level that might need further analysis, such as time of day, day of the week, age of causal driver, single vs. multiple vehicle involvement, causal vehicle type, general locations (e.g., county, city, rural, urban, roadway classification, locale, etc.);
- CMV higher severity crashes against CMV PDO crashes;
- CMV special target crashes (e.g., those caused by young drivers) against CMV crashes that did not involve the target (can be repeated for different severity levels if this is not already taken into consideration).

For the most part these comparative runs are designed to surface variables that have values with particular over-representations that point to the "who, what, where, when and why" of these crashes. Additional cross-tabulations and drill downs are recommended to get to the root causes of the differences that are uncovered at this high level.

Examples given in the Supplemental Report cover the following (section references are made to the Supplemental Report):

- Causal driver age for non-CMV causal drivers in CMV-car crashes (3.2.2.1);
- Point of initial impact for CMV-car crashes (3.2.2.2); and
- Time of day and day of the week for targeting (3.2.2.3).

These are given as examples, not necessarily recommendations, for the types of information that should be discovered prior to the implementation of the actual project. Each state and each project is different, and literally hundreds of variables were considered as part of the Alabama project problem identification to obtain as much information as possible to support the total effort.

3.2.3 Problem Identification Tools Available to the Patrol Officers

Web-based problem identification tools were developed to give patrol officers a direct hands-on capability to discover and study location information. In particular, the Supplemental Report details the following:

- Hotspot identification program,
- Details of crashes at a particular location,
- Hotspots represented as segments,
- Hotspots depicted on maps with imagery, and
- Patrol routing tools (PatrolSim).

3.3 Establish Control Areas and Study Times

3.3.1 Control Areas

A control area is defined as an area in which at least one component of the countermeasure group will not be implemented. Controls are necessary to assure that a variation in the before-after comparison for the test location is not due to some factor independent of the TACT program. Ideally a control area would be in all other aspects identical to the test area with the exception that one or more countermeasures are not implemented during the project duration. Since target (or test) locations are being selected at this point in the pre-project analyses, it would seem reasonable for this to be the best time to select control locations.

At times legal or ethical issues might arise in arbitrarily assigning a countermeasure to one area and not to another. This does not have to be the case if it is recognized that it is probably not optimal for the state to perform the TACT program statewide and consume all resource in a short period of time. Usually there are certain areas (perhaps those with the worse problems) that are to be given first consideration, and once that wave of the program is completed then projects are implemented in other areas. Careful planning and site selection at this point can create test and control areas without either of the locations being neglected – they are just covered at different times.

Perfect controls are not possible in the traffic safety arena as they might be in other fields (e.g., medical clinical trials). It is unusual to find two areas that are identical in population, traffic flow, CMV traffic density and any other confounding attributes. This should not deter project leaders from doing their best to obtain control locations that are as good as possible.

When rates (as opposed to raw numbers of crashes) are used in the before-after comparisons this provides an inherent control. For example, when CMV crashes/Total crashes (could be further qualified by severity or other factors) are used to provide the comparison metric, this automatically accounts for many external factors. A good application of this would be to the major reductions in crashes in the 2007 through 2009 calendar years cause by economic factors. If there is no reason to believe that CMV crashes would be affected any differently than non-CMV crashes, using total crashes as a denominator would create a perfect control to buffer out the economic effect. There is an argument that traffic in general might be affected more than CMV traffic, and if this is the case then further adjustment would be needed. If particular types of CMV crashes are being considered, then forming the ratio of the subtype to the total number of CMV crashes could create a metric that would buffer out the economic effects.

3.3.2 Study Times

The establishment of the times for the study is as critical as the determination of the test and control areas. It should obviously start with the "during" time for the project, since this should be established at this point (although it is always subject to change right up to and through implementation). Since crash rate and crash frequency are highly seasonal, especially in states with heavy snow and ice, it is best to choose comparable months for the comparisons. For example, if a project is to last three months, a three-month period from the previous three years could provide the before data. The data during the project itself would provide for the "during." The after should be a time span immediately after the project, and it might vary in order to determine the length of the "halo effect." (This is a term used in traffic safety literature to describe the continued deterrent effect that the observation of a law enforcement detail produces.)

It is important to give thought to just what is meant by "after." Theoretically it would be ideal if the "after" period were identical to the "before" period in number of months, and also in the particular months under consideration. For example, the first six months of the year should not be compared to the second six months because of seasonal variations; it would be better to compare the second six months of the previous year (before period) with the second six months of the current year (after period). If data are available, setting up the before period to be longer than the after period may be desirable for two reasons: (1) to get a more accurate reading of the "before" situation, and (2) to establish the variability of the before data for statistical purposes. It may also serve to establish any seasonal patterns or overall trends. And, of course, the longer the after period the longer the delay in getting the evaluation study accomplished.

If standard crash and citation data are employed, there is much more data available in the past than in the time since the project ended. Obviously sufficient time must be allowed for data compilation after the project to generate the data needed for the evaluation. It will be clear at this point just when the formal selective enforcement portion of the project ended. So, at this point the data can be partitioned into three time phases – before, during and after the project. The most definitive impact of the project would be expected for the "during" time period, and the effect will generally diminish after that unless there is some activity to sustain the effect. One important aspect of the evaluation will be to measure just how lasting the effect is.

In summary, reality dictates that the three time periods be viewed as follows:

- Before – an optimal amount of time before any of the effects of the project can be expected; the optimization balancing the increased accuracy and use of the before data to establish trends and variability against the decreased accuracy and thus misinformation that can be caused by obsolete data.
- During – the easiest time frame to establish since it should be well defined by the administrative data that has been collected.
- After – defined dynamically as more data is obtained. Initially it will be a combination of the data gathered during the project with a short time period after when the project may reasonably thought to still have an effect. As additional time goes by a second set of after data may be collected to determine at what point the residual effects (often called the "halo effect") of the project wore off.

No single metric is recommended to evaluate countermeasure effectiveness. Several metrics, including "before—during—after" studies with and without controls using raw numbers and rates, are recommended collectively to begin to create a picture of, not only the benefits being created by the program, but ways in which the program can be continuously improved.

This discussion of control areas and times has concentrated on crashes. It should be clear that if comparisons of citations, surveys or observational data are to be part of the effectiveness evaluation, a set of control areas and times would be required for them as well. A rationale similar to that used in defining study times and control areas for crash analysis should remain valid for other such evaluations should be valid.

3.4 Gather "Before" Crash Data in Target and Control Areas

The word *gather* is being used accommodatively in this and the next section to refer to the computer runs that will be necessary to summarize these data in a form that they can be used in the comparisons that are anticipated. The data has already been gathered in the form of crash records and they will be in the crash database according to the normally expected delay time. It is assumed that sufficient time has elapsed that the "before" period data will be available. If that is not the case, then computer runs can be made on the data that are available. This can form a type of "practice run," the intent being to assure that the data will be adequate for the intended purpose and according to the planned evaluation. These test runs will surface any deficiencies in the retrieval programs and perhaps in the data themselves.

3.5 Gather "Before" Citation Data in Target and Control Areas

If the evaluation is planned to entail citation data, it should be run at this time according to the above considerations to assure that all will be available at the appropriate time.

3.6 Gather "Before" Survey Data

The problem identification steps may result in a number of new facts that will need to be worked into the survey plans, and the data collection itself. This might include the following:

- Specifics for obtaining permissions to administer the surveys;
- Specific locations that will be targeted by the PI&E component of the project;
- The exact motto, target subpopulations and approaches; and
- The experimental design for the surveys, and in particular the sample size requirements.

Surveys may be administered to police officers, truckers or drivers in general. In order to get a baseline for determining any attitude changes, this survey data should be collected well before the TACT projects under consideration are conducted. A good rule of thumb would be a sufficient time to allow at least as many survey forms to be collected as is expected from the "during" and "after" surveys (individually). If the survey evaluation plan calls for control data to be collected, then it should be collected from the control areas defined for the crash and citation data above. This will require that the survey forms be finalized at this point; this will require an update of the preliminary design documented as part of the project plan.

3.7 Gather "Before" Observational Data

Observational data according to the plan, as discussed above, should be gathered at this time both to create necessary data for comparative purposes and to determine if any refinements in the data gathering process are necessary. False starts are common, and the complications of gathering data that has never been gathered before in a given way will inevitably result in unexpected changes to improve the process. For example, if video camera gathered events are to be analyzed, it will be important to resolve the definition of what constitutes an unsafe event. It will also be important to establish the "denominator," i.e., what qualifies any event to be included in the study. Will it include the mere observation of a CMV?, or will there need to be other vehicles in the CMV's vicinity in order for the event to be included? These details can get extremely complex and subject to the particular observers biases. It is very important that if multiple observers are used that they have frequent meetings and compare notes. Consistency in recording is one of the most important aspects of obtaining useful observational data.

3.8 Establish Administrative Data Support

The administrative data structure that has been established cannot be fully tested until the actual selective enforcement program is put into effect. It is assumed that this is not a new aspect of the agency or agencies that are administering the TACT program, since the maintenance of administrative data is a requirement for most government sponsored selective enforcement programs. Additional considerations for augmenting the normal administrative data collections were discussed in Section 3.1.6. It will be recalled that administrative data as to the time and place of implementation was not restricted to the selective enforcement effort, but it also extended to the

PI&E that is implemented as part of the TACT program. This will be essential for properly generating and interpreting crash, citation and survey data.

4.0 Data Gathering During the Project

The "project" here would include both the PI&E and the selective enforcement components. Generally the PI&E will be implemented slightly before the selective enforcement, depending on the strategy that was formulated in the planning steps. The most important data to capture during the project is administrative – it is essential to capture the exact start and completion times of each component. Without this data it may not be possible to pinpoint the crash and especially the citation data.

4.1 Feedback during the Project

One of the major reasons for doing some computer runs on crashes and citations during the project is to provide feedback to the project itself. Typically crash data will not be timely enough to provide information during the project. However, citation data are usually available in time to determine if some of the objectives are being accomplished, especially if the state has an electronic citation or a centralized citation intake system at the court level. Also, once the "before" surveys are put in place, there is no reason to stop collecting these data right through the project duration and into the "after" period. Early analysis of the surveys may provide feedback that can immediately improve the surveys.

4.2 Gather Crash, Citation, Survey and Observational Data during the Project

During the project much of the data may be limited. To the extent possible, preliminary data runs should be considered for any of the data gathering steps discussed in Section 5.3

4.3 Gather Administrative Data during the Project

This is the most important data to be gathered during the project, as discussed above. It should not be left to chance or solely to the individual officer. The collection of these data should be monitored to assure completeness and any incomplete or missing forms should be immediately resolved.

5.0 Data Gathering After the Project

It should be clear that consideration of data gathering is not something to defer until this point. If all of the steps above are followed, it will be almost trivial to generate the data needed to perform an effective evaluation.

5.1 Use of Data Gathered After the Project

There should be concern, not only with what the effects were at the actual implementation of all components of the project, but also any lasting effects. Viewing our goal from a cost-benefit point of view, the longer the halo effects continue the more benefit can be attributed to the project. Ideally, once a project is implemented its effect would continue forever. The goal here is not just to estimate this benefit, but also to determine the optimal implementation of a follow-up "wave" of the program. This adds a new dimension to the evaluation project – that of determining the size of the waves. Should all resources be expended in a single extremely large wave, or should they be dispersed over the duration of the funded program, say, for example, over five or more waves perhaps timed between traditional selective enforcement that concentrates on holidays. (Holidays might be the worse time to perform TACT projects since there are generally fewer CMVs on the road on or near holidays.) With these basic goals, multiple objectives of gathering data at a variety of time frames after the project include:

- To determine the total impact of the project;
- To determine at what point the project no longer has an impact (the duration of the halo effect);
- To determine at what point it would be best to stage a follow-up project – a decision that is as dependent on the availability of resources as anything else;
- Thus, to get a handle on the number and size of the project waves to be conducted in future programs.

5.2 Gather Crash, Citation, Survey and Observational Data after the Project

The following sections concentrate on the specific data elements to be gathered for each type of data. This chronological stepwise review of the methodology is in opposition to the thought process required for project planning. In other words, it is imperative that the statistical analysis be defined as a prerequisite to identifying/defining the data to be gathered. This was described in Section 3.1.3 and will not be repeated here.

The following are considerations for the collection (or generation) of data after the project:

- Crash data. Run the before, during and after time periods for all crashes, CMV crashes and CMV crashes that might be specially targeted by the project (e.g., CMV crashes that were caused by youthful drivers). Obtain this data for both the test and the control areas.
- Citation data. Run the before, during and after time periods for all citations, and obtain a breakdown by citation type for both the test and the control areas.
- Survey data. Although surveys might be continuing and data are continuing to be produced (as is obviously the case with crash and citation data), establish a cut-off point for the evaluation and compile the data in their before, during, and after time frames. Each question will have a proportion assigned for each of the responses.
- Observational data. This should now be available in terms of the total number of observations (e.g., the total number of CMVs observed or the total number of CMVs observed in which there were private vehicles in the vicinity), and the total number of these observations in which there were various types of faults on the part of one of the drivers. Each

fault should be broken out separately to determine if the project may have had an effect on one type of fault and not another. For example, tail-gating type faults should have been recorded separately from blind-spot faults.

In all cases above the raw numbers should be retained even though proportions are to be used in the analyses. This is because the sample size is required to determine the significance for statistical tests involving proportions.

5.3 Gather and Summarize Administrative Data after the Project

Generally this will be a requirement of the sponsoring agency for accountability purposes. It is urged, however, that the statistics gathered here not just be driven by these requirements, but that they also generate a permanent record of the data critical to the evaluations (e.g., the start and end times for the various project components, and the locations where they were implemented). This will enable future replication of the data runs should that become necessary.

6.0 Analytical Techniques (Statistical Analyses)

The statistical analyses that will be recommended in this section have the following objectives:

- They are simple and easy to understand;
- They can be implemented by traffic safety practitioners who have only had a basic statistics course or who have acquired a knowledge of statistics from reading a basic statistics textbook;
- They can be implemented using an Excel spreadsheet with the basic statistical "add-ins" or "formulas."

The recommended approach is in no way an attempt to short-circuit more sophisticated techniques, and their use by academics, consultants or anyone who might bring that expertise into the project is certainly encouraged. However, remember that the goal here is neither to prove the project's benefit nor to "tease out" some preconceived result. The purpose is to get an appreciation of evaluation metric differences between time periods and the test and control areas in order to assess the TACT project being conducted and continuously improve the process.

This section will continue by considering the various data types that should be available at this time.

6.1 Crash Data Analysis

Hypothetical examples that are given in this section will use theoretical crash data to illustrate the basic concepts. At the outset it should be recognized that these techniques are not at all limited to crash data, and that the same basic technique can be used on most of the data types (e.g., citation, survey, etc.) that are under consideration.

Data (raw numbers) typically do not produce information until they are compared to other raw numbers. Here the *raw numbers* are samples of reality. While they may be perfectly accurate and complete in stating what occurred in the past, it is impossible to state definitively that exactly the same will occur in the future, and thus to accurately predict what the effectiveness of a given future TACT project will be. Therefore, past experience must be viewed as a sampling of what is expected in the future given that the same implementation details are applied. Even the "before" data must be considered as a sample because the events (e.g., economic) that surrounded the capturing of that data will almost certainly never be exactly replicated again.

Statisticians handle these problems by using a variety of techniques that have been found to be quite effective in the past for making comparisons and forecasting future effectiveness of similar actions that have been studied experimentally. To understand these techniques it is essential to recognize that any given sample result is subject to a large number of in addition to the countermeasure(s) for which an assessment is being made. The more samples that are taken, the more reliable will be the *average* of these samples (also called the *expected number*) in producing a model of reality.

While the concept of averaging is well understood by most, the measure of the variation in the numbers is a bit more vague. There is a number that can be computed from any group of sample numbers called the *variance*, and closely related to it, the square root of the variance, which is called the *standard deviation*. These parameters measure the relative variation among the sampled numbers. Intuitively, sample numbers with a higher variance will be less reliable in predicting future events than sample numbers with a low variance (i.e., the ideal being a zero variance where the same sample number is observed over and over again).

Statistical methods use *averages* (also called *means* or *expected numbers*) and their corresponding *variances* (or *standard deviations*) to determine the probabilities that the differences between two sample averages are merely due to chance. Usually if this number is less than a given threshold (e.g., 0.05 or one in twenty), then they feel confident in concluding that the differences are *significant*, or more properly, *statistically significant* at the 0.05 level. There is nothing sacred about 0.05, and in some cases where a greater assurance is required, there might be an insistence on a 0.01 (one in 100) level of significance. In other cases where "being perfectly sure" is not nearly as critical, a one in ten (0.10) level of significance might be acceptable, or even less. *The major thing to recognize is that the level of statistical significance is a relative measure of confidence that the difference is not merely due to chance.*

The following sections explain some simple statistical tests that are used in the Supplemental Report for determining if observed differences are statistically significant. These will be presented in terms of examples. For a more theoretical treatment of the subject, readers are referred to any basic statistics textbook.

6.1.1 Student's t-Test Method (Using Excel)

Assume that a TACT program has been run statewide over a four month period in 2009, and the goal is to take the previous three years (2006-2008) and compare the number of CMV involved crashes to determine if there had been a statistically significant reduction in the overall number of CMV crashes. Further, assume that data for comparing the 12-months (Sep-Oct over three years 2006-2008) to the comparable "during" period (Sep-Oct 2009) is given below:

Before	During	
278	226	
289	233	
264	187	
256	201	
229		
292		
255		
209		
226		
233		
187		
201		
243.25	211.75	Averages
0.031		Probability of Difference due to Chance

The average number of crashes per month of this type in the before months was 243.25 crashes per month. The number observed in the after months was 211.75. This comparison found a difference in the means of 31.50, which was significant at the 0.031 level. The estimate of the number of this type of crashes that were reduced monthly during the term of the project (i.e., September through December, 2009) is 31.5 crashes per month.

Of interest here is the method for computing the probability that the difference between these two averages ("before" and "during") are merely due to chance. These raw numbers were placed in an Excel spreadsheet in the following rows and columns:

- D3-D14 for the 12 monthly "before" crash frequency numbers.
- E3-E6 for the four "during" crash frequency numbers.

The Excel function applied to obtain the Student's t-test probability was:

=TTEST(D3:D14,E3:E6,1,3)

where:

D3:D14,E3:E14 are the data ranges explained above, 1 = number of tails = the specification for a one-tailed test, and

3 = type = the type of test that has two samples with unequal variances.

Alternatives for number of tails. A two-tailed test would be used when the analysis is not concerned with which of the two (in the case the “before” and “during”) samples is the larger, only with whether they are different. Most traffic safety comparisons attempt to establish whether the “during” (or “after”) sample is strictly *less than* the “before” sample, and therefore a one-tailed test is most appropriate.

Alternatives for type. Excel provides for three alternative types, as follow:

1. Used for a “paired” t-test, when there are the same number of samples in the two sets of data being compared.
2. Used for two (generally unequally numbered) samples with the assumption that the variance of the two samples is equal.
3. Used for two (generally unequally numbered) samples with the assumption that the variance of the two samples is not equal.

Type specification 3 is used since there are unequal sample sizes and do not wish to make any assumption about the underlying population variances.

6.1.2 Binomial Probability Method (Using Excel)

There are times when it is more appropriate to compare some fractional probability of occurrence than simply raw sample numbers (as above). For example, the relative number of CMV crashes compared to crashes in general. Such a fraction can be viewed as a probability – in this case the probability of you arriving at any crash in general and finding that a CMV was involved. Another example would be the probability that a given citation involved tail-gating. Comparing these probabilities (fractions or ratios) over time might give an idea as to whether an emphasis on tail-gating is having an effect as far as the generation of citations is concerned.

The use of ratios as opposed to raw numbers has an effect of damping out the effect of other major causes. For example, if it is assumed that CMV and non-CMV traffic have both been affected essentially the same by economic factors, then the numerator and denominator of a fraction of CMV crashes/non-CMV crashes would be similarly affected by a recession. The fraction would be effective in determining whether CMV crashes have increased or decreased despite any changes in the economy. In a sense the denominator serves as an additional “control” in damping out those factors that might have nothing to do with the TACT program.

Consider a hypothetical TACT program that was implemented in the October to December 2009 time period. A decision is made to estimate the established proportion of crashes from the past three years of data, but since monthly variations could enter in, only the October to December time intervals in the past three years will be used. The metrics that will be examined in this example will be the values of the Primary Contributing Circumstance variable for CMV-related crashes. The following table illustrates the data generated from the crash records system. (Some crash data analysis systems have the ability to either generate the following directly or to export the data to a system like Excel where it can be manipulated into this configuration. All state crash data analysis systems have the ability to generate these data in some form.)

Hypothetical Example Summary of CMV Crashes
by Primary Contributing Circumstance Value

	Oct-Dec 06-08		Oct-Dec 2009		Odds	Prob 2009	
Value of Primary Contributing Circumstance	**Number**	**%**	**Number**	**%**	**Ratio**	**=> 06-08**	
Improper Lane Change/Use	150	12.50%	38	9.50%	0.76	0.0373	**
Improper Passing	36	3.00%	7	1.75%	0.58	0.0862	*
Improper or No Signal	48	4.00%	10	2.50%	0.63	0.0733	*
Over Speed Limit	21	1.75%	8	2.00%	1.14	0.7303	
Aggressive Operation	10	0.83%	8	2.00%	2.40	0.9929	
Driving too Fast for Conditions	30	2.50%	4	1.00%	0.40	0.0278	**
Followed too Close	75	6.25%	34	8.50%	1.36	0.9706	
SUBSET TOTALS	370	30.83%	109	27.25%	0.88	0.0659	*
GLOBAL TOTALS	**1221**		**402**				

* Significant at the 0.10 alpha level.

** Highly significant at the 0.05 alpha level.

There are dozens of codes that constitute the Primary Contributing Circumstance (PCC) attribute in most states. The ones listed above are those that were considered to be relevant to the TACT program. The following presents an explanation for each of the columns of the table given above:

- Oct-Dec 06-08 Number – the raw number of crashes (in this case CMV crashes) that had the respective number of the corresponding PCC in the Oct-Dec 06-08 time period (referred to as the "before" time period). Note that this is a total number over three years, and thus to make the Oct-Dec 2009 data point comparable to this it will be necessary to multiply the one-year number by three. Also, note that there are two "TOTALS" given at the bottom of the table for most of the columns. These will be described below.
- SUBSET TOTALS – This is the sum of the numbers above. It pertains only to the subset of codes that are included in the analysis. There is no reason to include all possible codes in the summary if only a few of them are relevant to the evaluation. In this hypothetical example the codes that are of interest are those that were, in the judgment of the analysts, relevant to the TACT program (i.e., involved aggressive driving).
- GLOBAL TOTALS – These totals pertain to all crashes that qualify regardless of the PCC code. For this hypothetical example, it represents all CMV crashes. Since the 1,232 for the before period is very close to three times the 406 for the after period, it can be

concluded that the total number of CMV crashes were about as expected. However, that factor is not relevant to this overall analysis as the primary purpose is to detect any significant changes in what is causing these crashes. The following items go back to discussing the headings on the table above.

- Oct-Dec 06-08 % – this is the percentage calculated by dividing the "Number" by the GLOBAL TOTALS number. Since the items listed are not for all possible codes, the SUBSET TOTALS for this column do not add up to 100%. It is important that these percentages be calculated over the GLOBAL TOTAL and not the SUBSET TOTALS. (If the subset totals were used this would calculate a conditional probability – conditioned on one of the listed codes occurring. The comparison would be valid, but only if interpreted in this way. Also, this would force some of the codes to have a positive odds ratio, which is not forced when the GLOBAL TOTALS is used for the denominator.) For brevity, this time period is called the "before" time period. This percentage represents that probability (or *odds*) that any CMV crash in the before time period will be caused by the corresponding PCC.
- Oct-Dec 2009 Number – the number of crashes corresponding to the Oct-Dec 06-08 Number (i.e., for that PCC value), but for the 2009 time period (in this case the "after" time period).
- Oct-Dec 2009 % – the percent corresponding to the Oct-Dec 06-08 %, but for the 2009 time period (in this case the "after" time period). All of the data to this point is obtained from the State crash analysis system. This percentage represents that probability (or *odds*) that any CMV crash in the after period will be caused by the corresponding PCC. Thus, while the values in the "Number" columns are not comparable (one was for three years, the other for just one year), the numbers under the "%" columns are comparable statistically.
- Odds Ratio – it was noted above that the % columns essentially provide an estimate of the probability of a given CMV crash being caused by the listed PCC. There are obviously two such estimates – one for the 2006-2008 time period and one for the 2009 time period. Although the raw numbers for the two time periods are not comparable (because the two time periods are of different duration), the percentages (or odds) are comparable. These will be compared to determine if there have been significant decreases in the after period. The odds ratio provides a preliminary indication of the differences. Obviously an odds ratio of 1.00 would indicate no change. Since this odds ratio is calculated as the after percent divided by the before percent, an odds ratio greater than one indicates an increase in the proportion, which is typically not desirable. An odds ratio of less than one indicates a reduction in the proportion, which is typically what the TACT program is striving for. The greater the deviation from 1.00, the greater the difference in the two proportions.
- Prob 2009 => 06-08 (in words, the probability that the proportion for the corresponding code is either the same of greater for the 2009 period as it was for the 06-08 time period)

– this is the probability that for this code things got no better in the after period than they were in the before period. For brevity, this is referred to as the "significance column." In statistical jargon this is called the *alpha* value or level, and it is interpreted as the probability of concluding that there is a difference (reduction) in the statistical parameter when in fact there is not. Thus, the smaller that this alpha value is the more confidence there is in concluding that there was a reduction in this particular PCC due to the TACT countermeasures that were applied.

- Significance indicator – as indicated by the footnotes, one asterisk (*) indicates that the number in the column to the left is less than or equal to 0.050, while two asterisks (**) indicate that the column to the left is less than 0.010. These indicators do not add any information, they are merely used here to make the statistically significant reductions easier to identify.

More discussion is warranted on the calculation of the probabilities in the significance column, and a number of different examples have been incorporated into this hypothetical example for illustrative purposes. The probability estimate given assumes a binomial distribution, which is described in any elementary statistics and probability textbook. The binomial distribution holds whenever there are only two outcomes for a given set of trials (e.g., when flipping a coin). In this case it is reasoned, for example, that the PCC is either "Improper Lane Change/Use" or not with the established probability given as 12.50% for any given CMV crash. This was established by the "before" three years of sampling and is thus the most reliable estimate that of past performance without a TACT program in effect. Similarly, the other PCC categories can be viewed in terms of their probabilities.

The hypothesis being tested is that *"there is no reduction in this probability in the after period."* The practical question is whether or not there was a reduction in one or more of the particular PCCs that relate to TACT programs, and if so, could that reduction be considered to be significant.

It is recommended that a standard statistical package be used to calculate these probabilities. For example, the above was produced directly from the =BINOM.DIST statistical function within Excel. This function returns the probability that a proportion derived from a sample does not deviate in a negative way

The arguments for this function are as follows (derived from the Excel help function):

BINOM.DIST(number_s, trials, probability_s, TRUE)

Where (example numbers from the "Improper Lane Change/Use" line of the above table):

- **Number_s** = the number of successes in the after trials (this is 38 in the example table above looking at the Oct-Dec 2009 Number" column of the "Improper land Change/Use" row). In this context a "success" is the crash being caused by this particular PCC.

- **Trials** = the number of independent trials (402, from the GLOBAL TOTALS row under the same Number column).
- **Probability_s** = the probability of success on each trial (12.50%, under Oct-Dec 06-08 % column).
- **TRUE** – this parameter is to direct the Excel function to generate the cumulative probability of Number_s *or less* successes. If set to FALSE, this function would return the probability that there are *exactly* Number_s successes, which is of no value for this application.

It might be asked at this point: how small does the significance column probability have to be in order for it to be considered statistically significant? *There is no rigid rule for this.* Obviously if a number less than 0.01 is obtained then there is less than one chance in a hundred that the proportion has not changed (been reduced) from the before to the after time period. So, if reductions in the probabilities are being sought (as is usually the case), then the smaller the significance column probability the better. However, even one in ten (0.10) or one in five (0.20) probabilities are not high enough to cause us to necessarily conclude that no change has taken place.

One reason that it is recommended that the raw numbers are given along with the percentages is so that the background, or the "evidence," for the conclusions are made clear. The term "evidence" is used because nothing involving statistics is ever totally proven. The objective is not to establish a claim, but to determine all that is needed to establish a true sense of reality upon which the most effective decisions can be based. The following can be observed from the table above:

- Improper Lane Change/Use – the numbers both before and after are relatively large and provide a reliable basis for drawing conclusions. The odds ratio is less than one so the proportions are going in the "right" direction. The probability that it is concluded that there was a reduction when in fact there was not a reduction in this PCC category is only 0.0373 – about one in 30 or so. Thus, there is confidence in concluding that there was a significant reduction in this metric.
- Improper Passing. The numbers here are considerably lower, so even though the odds ratio is lower than the category above it, the alpha probability (of making an error) is considerably higher (0.0862 as opposed to 0.0373). Nevertheless, it is still a relatively small probability and good sound evidence that there has been a reduction in this category.
- Improper or No Signal – this is quite similar to Improper Passing.
- Over Speed Limit. Note from the 1.14 odds ratio and a comparison of the proportions that the after metric is relatively larger than the before metric. There is never a way that a reduction can be inferred in such a case. On the other hand, there is no reason to conclude that there has been any significant increase in this PCC category.

- Aggressive Operation. This is a new category and it might not have being used in the past. External causes for dramatic changes in percentages should be explored. With these low numbers it would be impossible to draw conclusions in any event.
- Driving too Fast for Conditions. The before sample size of 30 or more is usually sufficient to put confidence in the before estimate. The fact that the after number is quite low should not be "held against it." For example, the ideal might be to reduce it to zero. The reduction from 2.50% to 1.00% under these conditions is strong evidence of a reduction.
- Following too close. It is quite clear that this metric has increased proportionally, and some cause might be sought for this difference since the large sample sizes both before and after give a quite reliable result. Of course, this significant increase should not be attributed to be caused by the TACT program. There is, however, always the possibility that the TACT program has made reporting officers more aware of a given PCC, and this would certainly be no exception.
- SUBSET TOTALS – this is the sum of all of the results from all of the individual categories, and it can be tested as well, since the categories that were chosen were those that would be affected by the TACT project. It shows a significant reduction, although not highly significant.

This completes the example. It must be noted, however, that these same principles can be applied to other crash variables. Examples might include:

- County or city – this would determine which particular jurisdictions within the project areas covered by the project were most affected.
- First (or Most) harmful event – this would determine if the crash type had significant changes.
- Manner of the Crash – this essentially provides a clue to the type of crash (e.g., whether tailgating or blind spots were involved). Another variable that gives a similar indication is "point of initial impact" for two-vehicle crashes.
- Vehicle type for the non-CMV vehicle.

In addition to the consideration of other variables, other subsets of data can be considered. For example, the above could be repeated for CMV crashes that are caused by young drivers. When identical analyses are performed on different subsets of data like this, the results for the different subsets should be compared to provide another dimension of insight. For example, a particular crash cause (e.g., often misjudging stopping distance) might produce significant result when applied to young drivers but not when applied to drivers in general.

Finally, it should be clear how control areas can be analyzed at this point. They are merely another subset. The same analyses that are done before and after on the test subset should be applied on the control subset and the results compared.

The section above may seem formidable from the point of view of having to perform a large number of analyses. It should be recognized that each comparative analysis that is performed will surface a different type of information and unveil something new that was previously hidden. There is no rigid methodology in performing the analyses, but it is highly recommended that broader categories (e.g., all CMV crashes) be analyzed first before analyzing smaller and smaller subsets. At some point the size of the subset might get so small that meaningful analyses are no longer possible. Given the analytical tools that most states have, it should not be difficult to conduct at least a dozen such analyses, and then to perform additional drill-downs when significant results are uncovered.

6.2 Citation Data Analysis

The factors that should be given consideration as far as citations are concerned are:

- The overall increase in citations – this can be determined by using the month variable and picking out the same months before and after the project for the comparison.
- The increase in certain citation types before and after – this is essentially the same as the model given above where citation type is substituted for PCC.
- A comparison of the test and control areas for the time of the duration of the project instead of comparing the before and after time periods. In lieu of a control area, the entire state exclusive of the test area can serve as a control.

These are the initial recommended analyses; as significant differences are uncovered it will lead to further analysis.

6.3 Survey Data Analysis

Most Internet-based survey packages have a statistical analysis component built in, and they are highly recommended. Some are provided free of charge if the survey is kept under a certain number of questions. However, the charges for these survey web sites are quite nominal and well worth the time and effort savings that they provide. The basic comparison for surveys is in the responses before, during and after the project. Demographics can provide for subset comparisons by repeating the basic analyses for each of the demographic types (e.g., gender, age, geographic areas, etc.)

6.4 Observational Data Analysis

As indicated in the discussion on the collection of observational data above, the number and proportion of the various types of driver faults will be available before, during and after the project. These can be analyzed directly by placing the results into the spreadsheet format given above.

7.0 Remarks on Drawing Conclusions

This manual, and the Supplemental Report, emphasize the value of viewing the results of the data analyses as *evidence* as opposed to absolute *proof*. Thinking in absolute terms tends to stand in the way of improvement. In fact, improvement is almost impossible without a healthy amount of self-criticism. Evaluations aimed at simply justifying projects and programs will tend to ignore many factors that might lead to improvement in these very efforts future. On the other hand, if the primary purpose of the evaluation is continual self-improvement, this removes the tendency to ignore or explain away discovered deficiencies.

7.1 Accomplishing the Evaluation Objectives

In Section 1.2 four objectives were given to for the evaluation of TACT programs. Now that the methodology has been established for the evaluation, how the output of these procedures can fulfill these objectives is discussed below:

- *To generally confirm the benefits of TACT programs and to establish the best estimate of its effectiveness in terms of reduced crash frequency and severity.* This is the traditional objective of evaluation, and it can be accomplished by a careful consideration of the data generated as a result of the objective evaluation of several TACT projects. These benefit estimates should not be made from the evaluation of one or just a few projects. Rather, they should accrue over time as evaluation becomes a normal part of the projects.
- *To find at least one weakness in each of the TACT components.* TACT *components* refer to the various selective enforcement components and the quite diverse PI&E components. The project evaluations should be organized and subdivided so that they are directed toward each of these components. The various evaluation components (e.g., crashes, citations, surveys and observations) should be designed to surface and address issues in the various TACT components.
- *To overcome these weaknesses by formulating recommendations for future TACT projects.* Those who are performing the evaluations should be objective in assessing and determining the strengths and weaknesses of the various aspects of the program. The identification of an issue is usually sufficient to establish a remedy for it. However, sometimes considerable creativity is required to remove the organizational and cultural obstacles that often stand in the way of improvement.
- *To seek out and establish, if possible, new and creative strategic approaches toward reducing the frequency and severity of CMV involved crashes.* This is a much more difficult edict that might call for the abandonment of some of the major assumptions employed in most TACT projects, and perhaps the adoption of approaches that are yet to be discovered.

7.2 Proper Use of Qualifiers

Even if a given project could be conclusively proven to produce a specific benefit, this would not mean that a subsequent implementation would replicate the benefit. There is no way to perfectly predict the future, and the CMV traffic environment is constantly changing. Similarly, if for some reason no benefit is measured, that does not mean that the project did not bring about a benefit (it could be, for example, that without the project the crash effects would have been significantly worse than in the before period). There are a host of factors that confound the ability to measure the benefits of traffic safety countermeasures. For example, recently a sharp increase in the price of fuel, in conjunction with the economic recession, had an especially large impact on young drivers and those in rural areas, resulting in a dramatic drop in the fatality number and rate. Unfortunately, it is to be expected that this process will reverse itself once the economy cycles back and these driver types increase their exposure.

Unqualified absolute statements with regard to the benefits of traffic safety countermeasures are highly suspect. This means that when a statement is made it should be qualified by a description of the situation in which the success (or failure) took place. The most relevant factors should be emphasized, such as:

- What were the conditions that existed prior to the program's implementation that made it particularly advantageous?
- What other programs were in effect prior to and during the project?
- When was the last time that a comparable effort was launched?
- What was the rationale for implementing the current program?

Again, the concept of building a case both for and against specific traffic safety activities is important. Traffic safety resources are scarce, and it is important to recognize that a downside of any project or program is the resources that it takes away from other (potentially more effective) efforts. This is the reason that the project decisions need to stay fluid and strive for continuous improvement forever.