SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites

Task K

White Paper for Module 1—Network Screening

For
Federal Highway Administration

GSA Contract No. GS-23F-0379K
Task No. DTFH61-01-F-00096

December 2002

SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites

Task K

White Paper for Module 1—Network Screening

Prepared by
Midwest Research Institute
iTRANS Consulting, Inc.
Human Factors North, Inc.
Ryerson Polytechnic University
Dr. Ezra Hauer

For
Federal Highway Administration
Office of Safety R&D
6300 Georgetown Pike
McLean, Virginia 22101-2296

Attn: Mr. Michael Griffith, HSR-20

GSA Contract No. GS-23F-0379K
Task No. DTFH61-01-F-00096

December 2002

Preface

This white paper presents a plan for developing functional specifications for *SafetyAnalyst* software Module 1, the network screening tool, and was prepared for the Federal Highway Administration (FHWA) under the requirements of GSA Contract No. GS-23F-0379K, Task No. DTFH61-01-F-00096. This paper, prepared as part of Task K, summarizes the technical approach to development of the *SafetyAnalyst* network screening tool. Preparers of this paper include Dr. Bhagwant Persaud and Mr. Craig Lyon, Ryerson Polytechnic University; Mr. Douglas W. Harwood, Dr. Darren J. Torbic, and Ms. Karin M. Bauer, Midwest Research Institute (MRI); Dr. Ezra Hauer; and other members of the *SafetyAnalyst* team.

MIDWEST RESEARCH INSTITUTE

Douglas W. Harwood
Principal Traffic Engineer

Approved:

Roger Starnes
Director
Applied Engineering Division

December 2002

Table of Contents

Appendices

Figures

Tables

SAFETYANALYST WHITE PAPER FOR MODULE 1—NETWORK SCREENING

Section 1 Introduction

In April 2001, the Federal Highway Administration (FHWA) entered into a contract with Midwest Research Institute (MRI) to plan and develop a set of software tools for safety management of specific highway sites, known as *SafetyAnalyst*. The *SafetyAnalyst* team also includes iTRANS Consulting, Inc.; Human Factors North, Inc.; Ryerson Polytechnic University; Woodward Communications, Inc.; and Dr. Ezra Hauer. *SafetyAnalyst* will incorporate computerized analytical tools that correspond to the main steps in highway safety management for site-specific improvements. The *SafetyAnalyst* team will first plan and then develop detailed functional specifications for the *SafetyAnalyst* software tools. Under a separate contract, computer software to implement each tool will be developed by a yet-to-be-selected FHWA contractor from the functional specifications.

This white paper, prepared as part of Task K, summarizes the technical approach to development of *SafetyAnalyst* software Module 1, the network screening tool. *SafetyAnalyst* will be comprised of four modules, which when packaged together, incorporate the six main steps for highway safety management:

Module 1. Network screening
Module 2. Diagnosis and countermeasure selection
Module 3. Economic appraisal and priority-ranking
Module 4. Evaluation

Thus, this white paper addresses the technical approach to the first of the four *SafetyAnalyst* modules.

This white paper expands on the material documented in the *SafetyAnalyst* Work Plan (May 2002) and incorporates changes in response to comments from the September 2002 meeting of the Technical Working Group (TWG), providing detailed explanations of the general approach to be followed during *SafetyAnalyst* development, and the actual network screening process to be implemented. It serves as a planning document as well as an overview of the module for the software development contractor. This will assure that the most appropriate network screening methodology will be incorporated in the functional specifications provided to the software development contractor. However, the approach that is documented makes several assumptions on the likely outcome of research on screening procedures that will be undertaken in this project and is currently underway in an associated project in Colorado.

This white paper provides a technical overview of the network screening tool, but it does not constitute a functional specification. The *SafetyAnalyst* team will develop a draft of a detailed functional specification for the network screening tool in Task L. That draft will be refined in response to FHWA and TWG comments in Task M. The detailed functional specification will then be provided to the software development contractor so that software development work can begin. It is expected that several key aspects of the network screening tool will remain to be resolved, even when the detailed functional specifications are complete. Thus, some pieces of logic may be incomplete in the functional specifications and some parameters may have values that are not yet quantified. Research to resolve these issues will be undertaken in Task E, in parallel with the software development effort.

This document is organized as follows. Following this introduction there is a general overview of the purpose of the network screening module. The third section discusses the planned capabilities of the module. The heart of the paper is the fourth section that describes the functionality of the main components of the module, specifically those in which inputs are requested, data needs and availability are assessed, required data are imported and processed, analyses are performed, and output is organized and delivered. In this fourth section, illustrations are provided that pertain primarily to road sections because it is felt that this type of entity requires the most sophisticated variations of the basic screening processes and methods. Applications to other entities, such as intersections and interchange ramps, are also discussed briefly and follow similar lines to those illustrated for road sections. The paper concludes with a section on planned development activities, followed by appendices that provide more details on data requirements and the technical approach to network screening.

Section 2
Purpose and Use of Network Screening Module

The basic purpose of the network screening module is to use available data to review the entire roadway network under the jurisdiction of a particular highway agency[1] and identify and prioritize those sites that have promise as sites for potential safety improvements and, therefore, merit further investigation, i.e., sites to which the other *SafetyAnalyst* modules should be applied. This network screening process will make use of information on roadway characteristics and safety performance to identify those sites that are the strongest candidates for further investigation. The data to be used fall under the following categories:

- Geometric design features
- Traffic control features
- Traffic volumes
- Accident history
- Accident characteristics
- Safety performance functions (SPFs)

The process of conducting detailed engineering studies of candidate improvement sites is an expensive one, even with the improvements in efficiency of such investigations that will be provided by the other *SafetyAnalyst* modules. Therefore, only a limited set of sites can be investigated by a highway agency in any one year. The most efficient network screening procedure is one that will best identify "sites with promise" as those sites (road sections, intersections, interchange ramps) that would most likely be the highest ranked in terms of safety cost-effectiveness among all candidate sites. This goal will govern the investigation and selection of practical approaches to network screening.

The basic function of the network screening module will be to rank sites by one or more selected measures or indices based on a consideration of each site's accident history, traffic volume, and roadway characteristics. However, the module will have other complementary capabilities. These are addressed in the next section.

[1] In this white paper, the term highway agency refers to both state and local agencies that are users of *SafetyAnalyst*.

Section 3
Capabilities Planned for the Network Screening Module

This section identifies the capabilities planned for the network screening module and also addresses some of the capabilities that the module will not provide. Details on how the module will function to achieve the planned capabilities are provided in later sections. Several of these capabilities will require further development and testing in subsequent tasks to determine the need for, and practicality of, implementing them. Users may need the option to customize the module and omit some capabilities if they do not have the data needed to support them. The need to balance ease of customization by an agency and ease of providing upgrades of the software will be considered during software development.

SafetyAnalyst, in general, will provide a range of capabilities based on data availability. Each tool, including the network screening tool, will provide a series of functionality levels, from basic to elaborate, to accommodate the user's needs and the completeness of available data. Clear data requirements and data dictionaries will be defined, refined, and updated as the work proceeds in developing the functional specifications for each tool's capabilities.

3.1 Planned Capabilities

It is expected that the following capabilities will be of most interest to a majority of users. The module will be able to:

1. Rank sites by appropriate measures or indices related to:
   - Potential for safety improvement (PSI) based on expected accident frequency
   - PSI based on excess accident frequency (amount by which the expected accident frequency exceeds that expected at similar sites)
   - Prospective cost-effectiveness based on expected accident frequency, excess accident frequency, or both
   - Overrepresentation of specific accident types (e.g., a higher than expected proportion of rear-end accidents at signalized intersection may indicate the need to adjust the intergreen period, adjust the cycle length, or implement some other accident countermeasure)
2. Provide flexibility and guidance for the user to choose among available measures/methods for ranking sites
3. Provide flexibility for the user to apply default SPFs provided with the software or to apply user-supplied SPFs

4. Rank sites separately, or in combination, by:
    - Type of roadway elements (e.g., roadway segments, intersections, interchange ramps)
    - Area type (rural/urban)
    - Terrain type (level/rolling/mountainous)
    - Geographic areas (entire jurisdiction, or specific regions, counties, cities, etc.)
5. Permit ranking based either on the sum, or weighted sum, of property-damage-only (PDO), nonfatal injury (NFI), and fatal injury (FI) accidents
6. Provide an option for the user to choose whether or not to rank by accident costs and to accommodate either default or user-supplied values for accident costs
7. Provide a geographic distribution of accidents within a roadway segment by accident severity level and identify points of concentration of accidents
8. Screening sites for specific accident types/countermeasures (e.g., run-off-road accidents for shoulder rumble strip or left-turn collisions for turn-lane installation)
9. Screening for sites that show deterioration in safety over time
10. Identification of "corridors with promise" through review of the safety performance of extended roadway sections
11. Screening based on a sliding-window approach for roadway segments

The screening of sites for specific accidents types/countermeasures will be of assistance to highway agencies in implementing targeted accident reduction programs like those anticipated in the implementation process presented in the NCHRP Project 17-18(3) guides.

SafetyAnalyst will be built on the concept of conducting screening based on *expected* accident frequencies. Expected accident frequencies can be estimated from safety performance functions (SPFs), which often take the form of negative binomial regression relationships to predict accident frequencies from traffic volumes and roadway characteristics. The Empirical Bayes (EB) method provides a means to combine SPFs predictions and observed accident frequencies into a single estimate of the expected accident frequency, so that the observed accident history of a site can be considered in the estimation process. The EB method used in *SafetyAnalyst* will be adapted from the approach currently being developed by the Colorado DOT. In addition, it is recommended that *SafetyAnalyst* include not only an EB approach to network screening based on the analysis of homogeneous roadway sections, as recently developed for the Colorado DOT, but also a traditional sliding-window approach to network screening for roadway sections that is updated to incorporate EB concepts.

3.2 Capabilities Not Planned for the Module

Conventional screening techniques utilizing observed accident frequencies and/or rates, often in a "statistical quality control" framework, are not currently being considered for application in *SafetyAnalyst*. Such techniques are known to have difficulties in identifying sites with potential for safety improvement because of the phenomenon of *regression to the mean*. Because of regression to the mean, sites with a randomly high crash count can be wrongly identified as being potential problem locations and sites that have documented problems may be missed due to a randomly low crash count. In addition, selection on the basis of accident rates tends to wrongly identify sites with low volumes because of the nonlinear relationship between accidents and traffic volumes. Appendix A provides additional discussion of these difficulties for the uninitiated reader.

It is recognized, based on a recent survey (Persaud, 2001), that several jurisdictions may currently be screening sites based on observed accident frequencies and/or rates, and therefore, for the purposes of continuity, it may be desirable for *SafetyAnalyst* to allow an option for screening based on these more conventional methods. However, it is also important that the process of identifying sites for possible improvements be as efficient as possible since resources can be wasted on sites that are incorrectly identified as potential problem locations and sites that truly have correctable safety problems may go untreated if not identified in the network screening process. Therefore, it is recommended that *SafetyAnalyst* not allow for screening based on observed accident frequencies and/or rates. As an alternative to performing network screening based on *observed* accident frequencies or rates, *SafetyAnalyst* will be built on the concept of conducting screening based on *expected* accident frequencies, which will be estimated as a weighted average of the observed accident frequency and the accident frequency predicted with an SPF.

3.3 Potential *SafetyAnalyst*/GIS Integration Capability

An interface between *SafetyAnalyst* and GIS would provide an attractive feature to the user to display sites on map displays, thematically color-code displays of ranked sites, and identify sites of interest by pointing and clicking. However, procedures for GIS integration have not yet been worked out and the feasibility of using *SafetyAnalyst* together with GIS data from multiple agencies has not yet been determined. Thus, potential GIS capabilities are not discussed in this document. Such procedures will be considered and worked out with the software development contractor when hired. At a minimum, it would be desirable for *SafetyAnalyst* to be able to input data with GIS location coordinates attached, to retain those coordinates during processing, and to output data with those coordinates attached for display.

Section 4
Functional Approach to the Module

This section outlines the anticipated functional approach to the network screening module in a level of detail that provides an overview of the logical framework for the software.

4.1 Overview of Basic Functionality

The following list outlines the basic functionality of the network screening module. More detail on each function is provided in subsequent sections. Basically the module will perform a sequence of steps as follows:

1. Request input options
2. Assess data needs and availability
3. Import and process required data
4. Perform analysis
5. Request output options
6. Produce output

It is also expected that the module will have a help function that will provide not only a "how to" but also background material related to the various analytical functions. Considerable resources will be required to develop the required raw material for this help function.

4.2 Input Options

The user will need to specify the following input to use the network screening module:

1. Type of screening to be performed (more explanation is provided in later sections)
   - Basic network screening
   - Detection of safety deterioration
   - Screening for overrepresentation
   - Screening for specific accident types
   - Identifying corridors with promise
   - Screening with sliding-window approach
2. Accident type and severity levels to be evaluated
   - Choose from some combination of severity (PDO, NFI, FI) and accident type [e.g., all accidents, accidents by number of vehicles involved, accidents

by impact type, object struck for single-vehicle accidents, and manner of collision for multiple-vehicle accidents, time of day (nighttime or daytime)]

3. Geographical area to be addressed
   - Statewide
   - Highway district
   - County
   - City
   - Sites of specific interest (**NOTE:** See Comment No. 1 below)
   - Other geographical descriptor
   - Any combination of the above
   - All of the above

4. Area type to be addressed
   - Urban
   - Rural
   - All of the above

5. Roadway elements to be addressed
   - Roadway segments (**NOTE:** See Comment No. 2 below)
     - Choose from two-lane/multilane, divided/undivided
   - Intersections (**NOTE:** See Comment No. 2 below)
     - Choose traffic control type, number of legs
   - Interchange ramps (**NOTE:** see Comment No. 3 below)
   - Railroad-highway grade crossings (**NOTE:** See Comment No. 4 below)
   - Any combination of the above
   - All of the above

6. Terrain type to be addressed
   - Level
   - Rolling
   - Mountainous
   - Any combination of the above
   - All of the above

7. Functional class of roadway to be addressed
   - Freeway
   - Arterial
   - Collector
   - Local (**NOTE:** See Comment No. 5 below).
   - Any combination of the above
   - All of the above

8. Jurisdiction to be addressed
   - Agency-maintained roadways
   - All roadways (**NOTE:** Need to consider data availability; see Comment No. 5 below)

9. Source of SPFs to be used (for each required SPF)

   - Default SPFs supplied with the software
   - User-supplied SPFs

All these inputs will define the network screening analysis to be performed and will be supplied by the user from input menus or dialogue boxes. The basic functionality of the network screening module is illustrated in Figure 1.

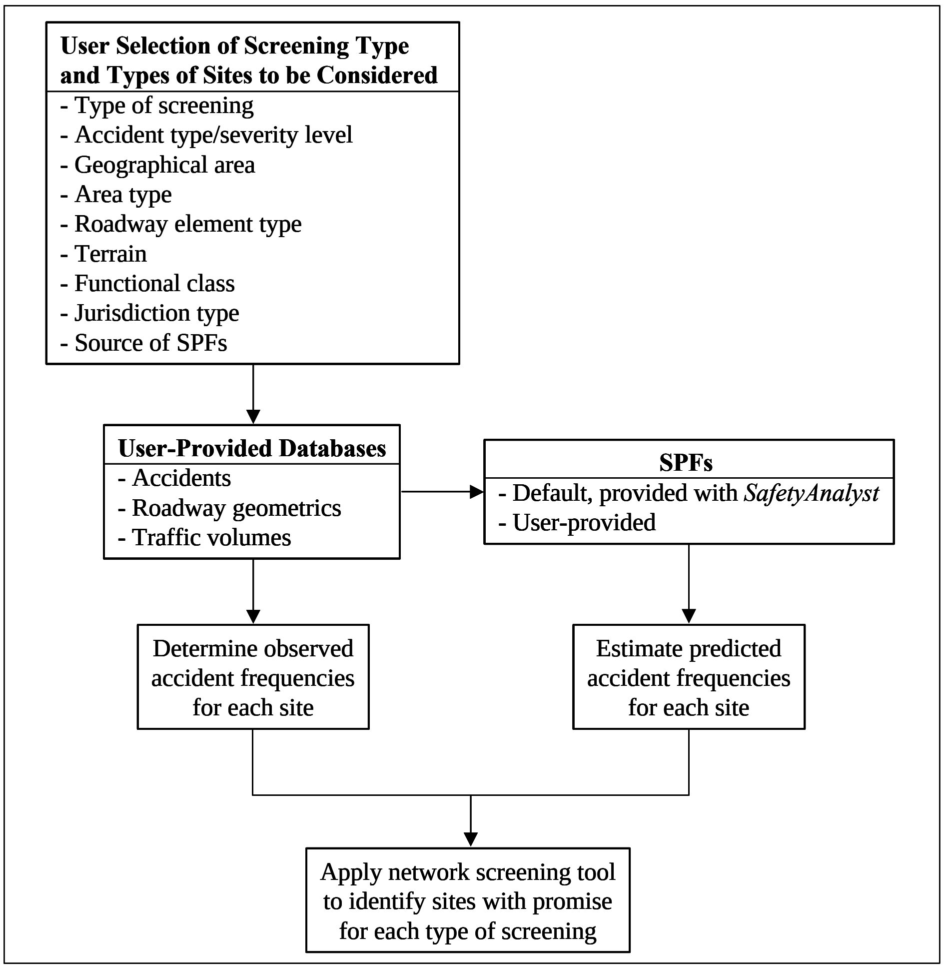

Figure 1. Overview of Module 1—Network Screening

Several issues identified above in the list of input data elements that control the scope of various network screening analyses need to be addressed.

> **Comment No. 1:** The basic idea is that for roadway projects that are already planned and, in particular, for those spanning long sections of roadway, it would be advantageous to be able to screen the project area for specific locations where a safety improvement might be implemented as a part of the planned project.

Implementing improvements in planned projects should improve the cost-effectiveness for the safety improvements since costs would likely be reduced.

The context of screening for sites of specific interest is somewhat different from conventional screening applications. One does not attempt to sift through a multitude of sites to select a few for more detailed examination. We are already at the detailed examination level, and the task is to spot opportunities for safety improvement at sites that are of interest for reasons other than safety. These are, for example, capital projects to improve level of service or pavement condition. The question is, "What will aid the user?" Certainly of interest would be a safety profile (a plot of estimated accident frequency and excess frequency with an indication of their precision) for any accident type the user might select. The Colorado project is making progress on this subject. Keeping track of those efforts with a view to implementing this capability in *SafetyAnalyst* is part of the planned developmental work in *SafetyAnalyst*.

Comment No. 2: It is clear that any useful network screening tool must consider both roadway segments and at-grade intersections. It is our recommendation that screening for roadway segments and intersections be conducted separately because they have different types of safety problems and require different SPFs. However, it is likely that screening for corridors with promise would address extended roadway segments and would include both intersection and nonintersection accidents within the corridor.

Comment No. 3: It would be highly desirable for the network screening tool to include the capability to consider interchange ramps; without this capability, the network screening tool would really not address interchanges at all. However, only a handful of state highway agencies have ramp inventory files and a consistent method of linking accidents to specific ramp locations. The inclusion of a screening capability for interchange ramps has been considered as an option and it is hoped that the availability of this option will encourage highway agencies to develop interchange ramp inventory databases.

Comment No. 4: It would be desirable to address the need for safety improvements at railroad-highway grade crossings. State highway agencies maintain a grade crossing inventory and have developed models to predict expected accident frequencies for grade crossings. However, this is typically done in a separate Federal program with its own mandated procedures and is conducted separately from the normal traffic accident records system. Because railroad-highway grade crossings are currently addressed in a separate program and a separate data set, we recommend that they not be addressed in the development of *SafetyAnalyst*. The capability to address railroad-highway grade crossings could be added to *SafetyAnalyst* at a later date if desired by *SafetyAnalyst* users.

Comment No. 5: The list of functional classes of roadways presented above includes the entire range of classes from freeways to local roads. In most states, the roadway system under the jurisdiction of the state highway agency includes freeways, arterials, and collectors, but does not typically include local roads. While it would obviously be desirable for *SafetyAnalyst* to address all public roads, including local roads, many states may choose to implement *SafetyAnalyst* only for the state-maintained roadway system. Availability and linking of geometric design, traffic volume, and accident data for locations off the state-maintained roadway system presents special problems. These issues will be addressed with the software development contractor when hired.

4.3 Data Availability and Needs

SafetyAnalyst will provide a range of capabilities based on data availability. Each tool, including the network screening tool, will provide a series of functionality levels, from basic to elaborate, to accommodate the user's needs and the completeness of available data. Clear data requirements and data dictionaries will be defined, refined, and updated as the work proceeds in developing the functional specifications for each tool's capabilities.

The requirements for undertaking the screening analysis consist of geometric design and traffic control information, traffic volume data, collision data, and SPFs. The module will need to access these items and assess their suitability for undertaking the requested tasks. There are two aspects to this function.

- **Handling Analysis Requests for Which Appropriate Data Are Not Available:** The *SafetyAnalyst* software will be aware of which types of data are and are not available in the databases supplied by the user. *SafetyAnalyst* will operate with several levels of capabilities depending upon the data needed. Only those tasks that can be performed with the available data will be presented to the user. If the user specifies an analysis that is not supported by the available data, that user will be informed that the task is not feasible with the available data; in addition, the data that would be needed to perform the requested analysis will be identified.
- **Determining Expected Accident Frequencies with SPFs:** A key element of all network screening procedures and many other *SafetyAnalyst* procedures will be estimation of the expected accident frequency for specific sites. Observed accident frequencies can only be judged properly through comparison to the expected accident frequency of a specific site. Expected accident frequencies in *SafetyAnalyst* will be determined from negative binomial regression relationships known as SPFs. *SafetyAnalyst* will be developed so that users will not need to develop SPFs themselves, unless they choose to do so. Default SPFs will be developed by the *SafetyAnalyst* research team and supplied to users with the *SafetyAnalyst* software. A calibration tool supplied with the software will allow users to calibrate the SPFs for particular site types and particular years of data.

No further interaction of users with the SPFs beyond this annual calibration process will be required. However, users that wish to develop their own SPFs will be able to do so and the software will allow the default SPF parameters to be replaced with user-supplied values.

4.4 Import and Prepare Required Data and SPF Information

This section outlines the raw data and SPF preparation processes envisioned for the module.

4.4.1 Importing and Use of Highway Agency Files

At this point in the development of *SafetyAnalyst,* the minimum and desired data requirements are not known and will not be until further development and testing of network screening methodologies and SPFs are complete. However, a preliminary list of desired data items is provided in Appendix B. These data items include, at a minimum, elements of traffic accident data, roadway segment, intersection, and interchange ramp data, cost data, and SPF data.

The data required for network screening will be stored in separate but linkable files. These files will be created from highway agency data files that will be imported into *SafetyAnalyst* and formatted appropriately to work with the *SafetyAnalyst* tools. *SafetyAnalyst's* utility/data maintenance package will include software tools to import data from highway agency files and convert the data to one or more standard formats that can be used by *SafetyAnalyst* tools. In particular, location data used to link files will need to be in one of several location reference systems that are selected for use with *SafetyAnalyst*. Preliminary plans for the utility/data maintenance package are presented in the *SafetyAnalyst* work plan. More detailed functional specifications remain to be developed. It is the research team's intention to make accident data definitions within *SafetyAnalyst* consistent with those in the Model Minimum Uniform Crash Criteria (MMUCC) to the maximum practical extent.

Based on the selected screening options, the tool will query all the required databases to develop a unique data file for screening and SPF assessment. The location identification information (e.g., route number and beginning and ending milepost of each segment) will be exported along with the available traffic and geometric information.

In the case of road sections, the module will have the capability to break down the data into segments of, say, 0.1 mi. The module will query the accident database and tally for each segment the number of PDO, injury, and fatal accidents occurring within its limits, linked by route number and milepost. The total number of accidents in each available year will then be added to the data file.

The basic roadway elements whose safety performance will be addressed by the network screening tool are homogeneous roadway segments of varying length, individual intersections, and individual interchange ramps. For any of these basic elements to be screened, its accident history, geometric design features, traffic control features, and traffic volumes must be available in the highway agency data files imported into *SafetyAnalyst*. As an alternative to the screening of homogeneous roadway sections of varying length, a screening tool using a traditional sliding-window concept is also being considered. This sliding-window algorithm would use an EB-based approach to estimating the safety performance of the roadway within the sliding window. The length of the sliding window would be user-specified.

4.4.2 Prepare Required SPFs

A database of default SPFs developed by the *SafetyAnalyst* research team will be provided with the software. The user will be presented with two options: apply the default SPFs built into the software or supply alternate SPFs. The choices made by the user when selecting from the list of input items (see Section 4.2) will determine the specific SPF needed to perform the screening process.

Option to use default SPFs: If the user does not provide any user-supplied SPFs, then the default SPFs provided with the software will be used. The software will be capable of deciding, for any given site with known characteristics, which of the available SPFs is most appropriate for use. Then, a simple lookup of the SPF and the determination of the appropriate SPF parameter values from the *SafetyAnalyst* database is all that is needed to estimate the expected accident frequency for the given site. It is envisioned that two levels of SPFs may be used:

- Level 1 (traffic SPFs): Accidents = function (traffic volumes)
- Level 2 (full SPFs): Accidents = function (traffic volumes, geometric design factors, other factors)

Level 1 SPFs will involve only traffic volumes as an independent variable. Level 2 SPFs will include traffic volume, geometric design features, and other factors as independent variables. Both SPF levels will be considered by the *SafetyAnalyst* research team in the modeling development process and whichever level proves to be most appropriate for a given model will be used. A decision has not yet been made concerning how SPFs will be used to estimate expected accident frequencies. One of two approaches will be considered: (1) use SPFs directly or (2) set some independent variables to nominal values and use accident modification factors (AMFs) to adjust for the effects of geometric design and other factors. The latter approach, using AMFs, was used in the crash prediction module of the FHWA Interactive Highway Safety Design Model (IHSDM).

Option for user-supplied SPFs: If a jurisdiction has estimated and is occasionally updating its own SPFs then, at the time the software is “setup,” an option will be provided to enter these SPFs into *SafetyAnalyst*. Any user-specified SPFs that are available will be used in place of the default SPFs provided with the software. Preliminary guidance for development of user-supplied SPFs is presented in Appendix C.

SPF calibration: It is envisioned that SPFs will need to be calibrated whenever new information becomes available, most probably on a yearly basis. Preliminary procedures for calibrating SPFs are presented in Appendix C. These calibration procedures will be implemented in a calibration tool to be supplied as part of the *SafetyAnalyst* software. This tool will allow users to calibrate particular SPFs and to review and accept the calibration factors developed using software with the *SafetyAnalyst* package.

Because the calibration process will account for year-to-year variations, the SPFs themselves should need updating much less often. Updates at five-year intervals would probably be sufficient. It is planned that FHWA would see that the SPFs are updated at appropriate intervals and would distribute the updated SPFs as part of routine software updates. User-supplied SPFs would be updated at intervals determined by the user. Implementation of updated SPFs will require calibration of those SPFs for all years of data to which they will be applied.

4.5 Performing Basic Network Screening Calculations and Manipulations

This section outlines the current vision of analytical methods to be performed by the network screening module. Covered are basic screening calculations. These functions include the Empirical Bayes (EB) estimation of the expected accident frequency for each site being screened, the calculation of the potential for safety improvement, cost-effectiveness calculations, and ranking of sites. A number of other potential applications are covered in Section 4.6.

Two basic approaches will be made available in *SafetyAnalyst* to identify sites with promise:

- a homogeneous section approach for screening roadway segments that can also be adapted to screening intersections and interchange ramps
- a sliding-window approach adapted from current highway agency procedures but enhanced by the use of SPFs and EB procedures and applicable to roadway segments only

The homogeneous section approach will be adapted from the procedures currently being developed by the Colorado DOT, using SPFs and EB procedures. A sliding-window approach, closer to current highway agency procedures, will be developed as well. This sliding-window approach can be applied to nonhomogeneous roadway sections and will be incorporated as an option in *SafetyAnalyst*. Enhancements to this approach will be added, however, to incorporate the use of SPFs and EB procedures so that the sliding-window approach incorporates scientific procedures for treatment of estimated and observed accident frequencies.

The remainder of this section presents the basic network screening approach for homogeneous sections. A hypothetical road section is used to illustrate the calculations. Sample calculations for a roadway segment are provided since they involve more complex issues than for other types of entities. For intersections or other location types such as ramps, for which the unit of analysis is clearly defined, the methodology would be the same except for the consideration of section length in any equations and the issue of what constitutes a "site." The sliding-window approach to screening is presented in Section 4.6.5.

4.5.1 EB Estimation of the Expected Accident Frequency

A road section consists of multiple subsections or segments of varying length. The homogeneous section approach treats each of the segments, whatever its length, independently and assesses whether the safety performance of that segment is of sufficient concern to be selected for detailed engineering studies. The homogeneous section approach considers the safety performance of fixed-length segments within the homogeneous section and also includes a "peak searching" algorithm to identify the segments with highest accident frequency within a homogeneous section. In the network screening methodology currently being developed by Hauer for the Colorado DOT, the analysis is based on roadway segments that are 0.1-mi in length. Because of the precedent set by the Colorado work, it may be desirable to use a 0.1-mi segment length in *SafetyAnalyst*. The network screening procedure will be tested to determine whether the 0.1-mi segment length is appropriate and a final decision will be made based on these results. Also, investigations of the compatibility of 0.1-mi segments with the records systems of various state-highway agencies are planned. The following discussion uses the 0.1-mi segment length for illustrative purposes.

The following example presents an application of the EB methodology. Each subsection or roadway segment is 0.1 mi in length except for the last one, which is the remainder. For each road section by class and terrain, there are two model equations: one for *Total Accidents* and the other for *Injury Accidents*. In addition, for each 0.1-mi subsection of roadway, we have its accident history in terms of PDO, NFI, and FI accidents. The accident history is available since 1989. We also know the date of the last major reconstruction. The accident history since 1989 or the year after the last major reconstruction is to be combined with the appropriate model equation to produce *three expected accident profiles* for PDO, NFI, and FI accidents. We will first produce model predictions for Total (PDO+NFI+FI) and Injury (NFI+FI) accidents. The model prediction for PDO accidents will be the difference of these two. We will have computed the ratio of FI/(NFI+FI) for each road and terrain class and will use it to produce separate models for NFI and FI accidents. These will then be used to estimate model predictions for NFI and FI accident frequencies.

The following demonstrates the EB calculations of the expected accident frequencies using a slightly simplified notation from Hauer. The terms used are defined below.

Term	Definition
y	• Subscript to represent the year • The first year for which data are available or the first year after major reconstruction is Year 1, i.e., y=1 • The last year, the year for which the ranking is produced, is year Y, i.e., y=Y
K_y	• *Observed* number of accidents on this road section in year y
$E(K_y)$	• *Predicted* number of accidents on this road section for year y using an SPF
C_y	• Yearly correction factor for year y relative to Year 1: $C_y=E(K_y)/E(K_1)$
w	• A "weighting" factor given to the accidents expected on similar entities
k	• Overdispersion parameter corresponding to the SPF used • For roadway segments, if the regression provided an estimate, k', per unit length, then k is adjusted for segment length as: $k = k' \times$ appropriate length • No adjustment to k for unit length is necessary at intersections
$\hat{K}_y$	• Weighted estimated number of accidents *expected* on this road section in year y, considering both the SPF estimate, $E(K_y)$, and the observed accident count, K_y, in that year
$VAR(\hat{K}_y)$	• Variance of $\hat{K}_y$

The objective is to calculate $\hat{K}_y$ and $VAR(\hat{K}_y)$ for subsections 1, 2,…,L, for each year and for each severity level (PDO, NFI, and FI accidents). To obtain these, proceed as follows.

Step 1: Using the SPF model parameters, compute, for each year, y, the $E(K_y)$ per unit length (say, 1 mi) for Total accidents and for NFI and FI accidents. Their difference is the model estimate for PDO accidents. Compute the ratio of fatal injury accidents to all injury accidents ($R_{FI/NFI\&FI}$). Multiply the model prediction for NFI and FI accidents by (1- $R_{FI/NFI\&FI}$) to obtain the model prediction for NFI accidents and by $R_{FI/NFI\&FI}$ to obtain the model prediction for FI accidents

Step 2: Using the model predictions for the three accident severities computed in Step 1, compute the three vectors of C_y for years y =1, 2,…,Y.

Step 3: Using $E(K_1)$,…, $E(K_Y)$ and the overdispersion parameter, k, for NFI and FI accidents, compute the w's, one for each accident severity class, using the equation:

$$w = \frac{1}{1 + k \sum_{y=1}^{Y} E(K_y)} \quad (1)$$

Step 4: Calculate the base $\hat{K}_1$ as:

$$\hat{K}_1 = wE(K_1) + (1-w)\sum_{y=1}^{Y} K_y / \sum_{y=1}^{Y} C_y \quad (2)$$

Step 5: Using the result from Equation (2), calculate $\hat{K}_y$ for *y=2,…,Y*, as:

$$\hat{K}_y = \hat{K}_1 C_y \quad (3)$$

Step 6: To obtain a measure of the precision of the expected accident frequencies calculated in Equations (2) and (3), calculate the variance of $\hat{K}_y$ as:

if $y = 1$:

$$VAR(\hat{K}_1) = \frac{\hat{K}_1(1-w)}{\sum_{y=1}^{Y} C_y} \quad (4)$$

if $y > 1$:

$$VAR(\hat{K}_y) = \hat{K}_y (1-w) \frac{C_y}{\sum_{y=1}^{Y} C_y} \quad (5)$$

4.5.2 Calculation of Basic Screening Measures

During the course of the *SafetyAnalyst* research, it is likely that a definitive screening methodology will emerge. For the purposes of this white paper, a number of possible alternate methodologies are presented. Fundamentally, these methodologies vary in two aspects, according to:

1. whether the potential for safety improvement is based on a site's expected number of accidents as calculated above or its expected excess accident

frequency (the difference between the expected number of accidents at the site and at similar sites);

2. whether the ranking is based on the two measures mentioned above or the prospective cost-effectiveness of potential safety improvements. This allows for different location types (e.g., four-legged intersection, two-lane undivided highway, interchange) and environments (urban, rural) to be ranked simultaneously given that average project costs vary by environment and location type.

4.5.2.1 Potential for Safety Improvement (PSI) Calculations

The PSI estimates how much the long-term accident frequency at a site could be reduced. There are two possible concepts:

1. PSI based on a site's expected accident frequency
2. PSI based on a site's expected *excess* accident frequency

Appendix D provides example PSI calculations for a rural two-lane road segment with a 13-year accident history based on that site's expected accident frequencies.

Similarly, Appendix E provides example PSI calculations for the same rural two-lane road segment based on that site's expected excess accident frequencies.

4.5.2.2 Prospective Cost-Effectiveness Calculations

If sites are to be ranked by prospective cost-effectiveness of potential safety improvements, then the expected project costs and benefits must be considered. The prospective effectiveness can be based on either the expected accident frequency or the expected excess accident frequency. The estimation of both of these is illustrated in Appendices E and F.

At this stage of the safety management process, the costs and benefits of projects are, of course, unknown. The actual costs and benefits of proposed projects cannot be determined until the *SafetyAnalyst* diagnostic and countermeasure selection tools are applied to help the user select a particular improvement type and the *SafetyAnalyst* economic appraisal tool is then applied to estimate the costs and benefits of that improvement. Since actual costs and benefits are not known. At this stage, estimated costs and benefits must be used. In work for the Colorado DOT, Hauer has proposed that prospective cost-effectiveness be based on average costs of particular project types and that benefits are proportional to either expected accident frequency or expected excess accident frequency. It has been proposed that average project costs should be estimated for the most disaggregate categorization possible. For example, average project costs may be classified according to area type (urban/rural), intersection or roadway type, traffic levels, etc.

Suppose the average annualized project cost, $Cost_{avg}$, for an urban 4-legged intersection with similar traffic volumes is $50,000, the *excess* expected accident frequency is 23.50 per year, and the expected accident frequency is 47.36 per year. The prospective cost-effectiveness measure based on *excess* expected accident frequency can then be calculated as:

$$Cost_{avg}/[\hat{K} - E(K)] = 50{,}000/23.50 = \$2{,}128/acc$$

and the prospective cost-effectiveness measure based on the expected accident frequency as:

$$Cost_{avg}/(\hat{K}) = 50{,}000/47.36 = \$1{,}056/acc$$

Sites would be ranked in ascending order of these numbers.

A decision will have to be reached, in consultation with FHWA and the TWG, whether prospective cost-effectiveness based on average project costs is sufficiently meaningful for use in *SafetyAnalyst*. It might be feasible, for example, to determine average project costs for projects in urban and rural areas; urban projects are undoubtedly more costly, on the average, than rural projects, but the range of costs for urban and rural projects is broad and overlapping.

4.5.2.3 Ranking of Sites

All sites meeting the user-selected criteria (e.g., geographical, area type, roadway element, terrain type, etc.) will be ranked according to the method selected by the user. The top ranked sites will be selected by the user for a detailed safety investigation that will be conducted using the other *SafetyAnalyst* tools. At this stage, the user should have the option to selectively remove any sites from the list that may not be suitable candidates for further investigation for any reason, and the list will be updated accordingly.

The ranking of sites for network screening purposes has, so far, been described as a process that considers road sections of varying length that are divided into 0.1-mi subsections. The following discussion addresses how “peaks” in expected accidents can be identified within a road section. These peaks can be considered in ranking the sites. Specifically, in the ongoing Colorado project, Hauer et al. (2002b) have developed a procedure that identifies, within a particular road section, the segment that has the highest potential for safety improvement while meeting requirements for statistical precision as outlined below.

The implementation of this procedure requires a database where accidents can be allocated to subsections within each road section. With such a database, the road section is divided into 0.1-mi basic subsections as shown in Figure 2. A window consisting of W consecutive basic subsections is said to be of size W. Initially, the left edge of the window is placed at the left boundary of the road section, and the average expected accident frequency within the window is computed. The window is then moved one basic subsection to the right, and the average expected accident frequency is computed again. This is done until the right edge of the window reaches the right boundary of the road section. The process is repeated for windows of all feasible sizes. The largest of the averages so computed is the largest peak for a window of size W. Segment AA' in the figure is the highest peak when W = 3. Segment BB' is the second highest peak when W = 3. Segment CC' is the highest peak when W = 7. If two segments of size W overlap, only that with the higher estimated average is retained for further consideration. For example, if on segment AA' the statistical precision criterion is met, segment AA' rather than segment CC' will be considered further. The reason is that the rank of segment AA' is bound to be higher than the rank of CC' and, if a diagnosis is conducted, the vicinity of AA' will be considered when a project is formulated. If the statistical precision criteria on AA' are not met but they are met on CC', only the latter will be retained for ranking.

To specify the required statistical precision, use of "limiting coefficients of variation" is suggested. The coefficient of variation of an estimate, denoted as "*CV*," is given by:

$$CV = \text{(standard error of estimate) / (expected estimate)} \qquad \textbf{(6)}$$

Let CV_a be the limiting coefficient of variation when ranking is based on assumption "*a*" and CV_b the corresponding value when ranking is by assumption "*b*."

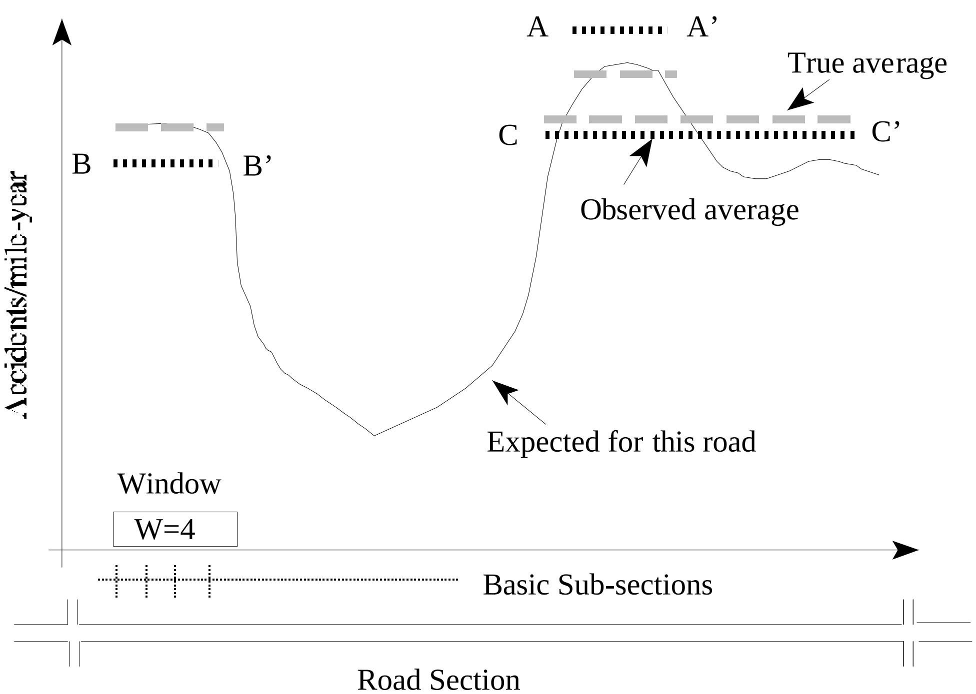

Figure 2. Illustration of the "Peak Searching" Algorithm

Then only sites for which the ratio:

$$\text{(standard error of estimate}_a\text{) / (estimate}_a\text{)} < CV_a \qquad \textbf{(7)}$$

and sites for which the ratio:

$$\text{(standard error of estimate}_b\text{)/(estimate}_b\text{)} < CV_b \qquad \textbf{(8)}$$

will be placed on the lists. Values of CV_a and CV_b are to be set by default or chosen by the user. The following can serve as guidance. The estimated value is usually within $100\times2\times CV$% of the true value and almost always within $100\times3\times CV$%. Thus, if a limiting CV of 0.033 is chosen, the estimated value is usually within 7% and almost always within 10% of the target.

The role of this "peak searching" algorithm in network screening will be evaluated further. A decision will have to be reached as to whether "peak searching" should be done in all cases or only when requested by the user. A decision also will have to be reached as to whether the CV_a and CV_b parameters should be set by the user or whether these should be set as default values that the knowledgeable user can change. A preliminary decision has been made to provide the user the capability to specify parameters of the peak searching algorithm.

4.5.3 Differences in Basic Screening Procedures Between Roadway Segments and Intersections

The previous sections focus on the basic screening methodologies for roadway segments because roadway segments require the most sophisticated variations of the basic screening processes and methods. Application to other entities, such as intersections, follow similar lines to those illustrated for roadway segments sections with some key differences. This section highlights the differences in the basic network screening procedures between roadway segments and intersections.

A fundamental difference between network screening for roadway sections and intersections is that intersections are discrete entities. Thus, intersections present no issues analogues to how to subdivide segments, how to search for peaks in accident experience within segments, and how the overdispersion parameter of the SPF varies with segment length.

The first key difference is that roadway segments and intersections require different SPFs. As illustrated in Section 4.4.2, SPFs for roadway segments will likely be provided at two levels of detail. For intersections, it is anticipated that SPFs will also be provided at these two levels of detail, but the major difference will be that the SPFs will likely incorporate two traffic volumes, that of the major road and that of the minor (or intersecting) road.

- Level 1 (traffic SPFs): Accidents = function (major-road traffic volume, minor-road traffic volume)
- Level 2 (full SPFs): Accidents = function (major-road traffic volume, minor-road traffic volume, geometrics, other factors)

Therefore, in developing SPFs for intersections, an agency's availability of minor-road traffic volumes needs to be considered.

The second key difference is that for roadway segments, SPFs are used to estimate an expected accident frequency per year per-unit-length. For intersections, SPFs are used to estimate an expected accident frequency, but the result is not proportional to any particular length. This difference in "dimensions" is accounted for when combining the information contained in the accident frequencies with the information contained in knowing the safety of similar entities. In the EB procedure discussed in Section 4.5.1, the expected accident frequency of an entity, whether it is for a roadway segment or an intersection, is estimated as follows (Hauer et al., 2002a):

Estimate of Expected Accidents for an Entity =
w × Accidents expected on similar entities + (1 – w) × Observed accident frequency on this entity **(9)**

where:

$$0 \leq w \leq 1$$

Recall from Equation (1) that *w* is calculated as a function of *k*, the overdispersion parameter of the regression model used as the SPF. For roadway segments, the estimated overdispersion parameter, *k'*, applies to a unit length; in that case, the overdispersion parameter, *k*, is adjusted for segment length such that *k* = *k'* x appropriate length. For intersections, length does not apply, so the unit value of *k* applies to all intersections in the appropriate class.

An important issue to be addressed is how to distinguish between roadway-section-related and intersection-related accidents. Most traffic accident records systems identify individual accidents as "at intersection," "intersection-related, but not at intersection," and "not intersection-related." Most accidents also have a milepost or coordinate value that indicates the location relative to roadway features. In previous research (Harwood, et al., 2000; Harwood et al., 2002), intersection accidents have been defined to include all at-intersection accidents and all intersection-related accidents within 250 ft of the intersection (or halfway between intersections if intersections are spaced less than 500 ft apart). We are aware of at least one state that assigns all at-intersection and intersection-related accidents to the milepost of the intersection. Some states also use a link-node system, where intersection-related accidents are assigned to a node number and roadway section accidents are assigned to a link defined by the two node numbers at the ends of the link. State records systems also differ in the availability/inclusion of accidents on the minor road at an intersection if the minor road is not a state highway. A review of the treatment of intersection and nonintersection accidents will be undertaken to assure that the network screening algorithm can work effectively with a variety of data types.

4.5.4 Differences in Basic Screening Procedures Between Roadway Segments and Interchange Ramps

The basic screening procedures for interchange ramps will be very similar to those for intersections. While a ramp is a roadway segment of varying lengths, ramps are relatively short and in many respects can be best treated as a discrete element rather than an extended segment. Both the ramp length and its traffic volume will need to be considered in estimating its expected accident frequency. In screening procedures, ramps will need to be classified by ramp type: diamond, loop, direct connection, etc.

4.6 Performing Other Network Screening Applications

It is envisioned that the network screening module will have a variety of additional capabilities that enhance the basic network screening capability described above. At this stage, these are relatively new procedures, so in some cases, only bare-bone details are available. These procedures will be developed more completely through Task E research. In addition, it remains to be determined the extent to which some of these applications are of interest to potential users. The need for each of these procedures will be discussed with the TWG. The additional applications covered in this section are:

- Detection of safety deterioration over time
- Screening for high proportions of specific target accident types
- Screening for high frequencies of specific target accident types
- Identification of “corridors with promise”
- Use of sliding-window approach for screening roadway segments

4.6.1 Detection of Safety Deterioration Over Time

A complementary methodology for identifying sites for investigation is to identify those that experience a gradual or sudden increase in mean accident frequency (Hauer, 1996a; and Hauer, 1996b). To illustrate the methodology, consider the following example. On a section of highway, the following accidents have been recorded over the past 5 years (Table 1):

Table 1. Recorded Accident History

Year_i	Recorded accidents, x_i
1	7
2	5
3	10
4	15
5	13

Time periods are numbered *1, 2,....,T, T+1,..., L*. The number of observed accidents in each time period is denoted $x_1, x_2,....., x_L$. For the end of any time period *T*, $(1 \le T < L)$, the difference between the "after" and "before" period accident averages is calculated as:

$$\Delta(T) = \frac{\sum_{T+1}^{L} x_i}{L-T} - \frac{\sum_{1}^{T} x_i}{T} \qquad (10)$$

For the example road section, the results are given in Table 2.

Table 2. Average Accidents During Before and After Periods

T	Recorded accidents, x_i	Accident average before	Accident average after	$\Delta(T)$
1	7	7.0	10.8	–3.8
2	5	6.0	12.7	–6.7
3	10	7.3	14.0	–6.7
4	15	9.3	13.0	–3.8
5	13	NA	NA	NA

Based on these results, two tests will be conducted. The first test is to detect a potential steadily increasing trend in mean accident frequencies. The second test is to detect a potential sudden jump in the mean accident frequency. The computations are illustrated next.

4.6.1.1 Detection of Steadily Increasing Mean Accident Frequencies

The objective of this test is to find the time period, *T**, at which the variability of $\Delta(T)$ is at a minimum. The method is as follows.

Let:

a = the mean accident frequency in period 1
b = the growth rate per period
L = the total number of periods of accident data

Where the mean accident rate is increasing at a constant rate, the expected difference in accident means for any time period *T* is:

$$E(Ä) = bL/2 \qquad \textbf{(11)}$$

and the variance of this estimate at time period *T* is:

$$VAR[Ä(T)] = L(a+bT)/[T(L-T)] \qquad \textbf{(12)}$$

If *b*>0, i.e., the mean accident frequency is increasing over time, then *VAR[Ä(T)]* is smallest at time period T* where:

$$T^*=(a/b)[(1+bL/a)^{0.5}-1] \qquad \textbf{(13)}$$

These calculations are now illustrated for the example highway section in Table 1.

Step 1: Calculate *Ä(T)* for all possible values of *T*, *T = 1, 2,…, L-1*. The results are shown in the last column of Table 2.

Step 2: Estimate *a* and *b* using simple linear regression to fit a straight line to the pairs of data shown in Table 1. This can be easily done in Excel, for example.

In the example, *a* = 3.4 and *b* = 2.2.

Step 3: Using Equation (13) and the values of *a* and *b*, calculate the time period, *T**, for which $VAR[\Delta(T)]$ is at a minimum:

$$T^*=(3.4/2.2)[(1+2.2(5)/3.4)^{0.5}-1] =1.64$$

Look up the value of $\Delta(T^*)$ in Table 2. In this example, $\Delta(T^*) = \Delta(2) = –6.7$.

Step 4: If *b* is larger than some predetermined value that takes into account the expected change in accident frequency due to changes in traffic and other factors, then *T** is calculated to the nearest integer value. Thus, *T* = 2.*

Step 5: Subject *Ä(T**) to a statistical test of significance and rank sites that pass. Sites meeting the statistical test for a gradual increase in mean accident rate can then be ranked as desired. A Basic computer program called HYPTEST2 written by Hauer is available for performing the necessary calculations. With some modifications, this program could be incorporated into the *SafetyAnalyst* software. Details of the HYPTEST2 program are documented in Hauer (1996).

4.6.1.2 Detection of a Sudden Increase in Mean Accident Frequencies

Screening procedures will be developed to detect a sudden increase in mean accident frequencies over time. Sites meeting the statistical test for a sudden increase in mean accident rate can then be ranked as desired. A Basic computer program called HYPTEST2 written by Hauer is available as an example of how the appropriate calculations can be performed. With some modifications, this approach can be incorporated into the *SafetyAnalyst* software. Details of the HYPTEST2 program are documented in Hauer (1996).

4.6.2 Screening for High Proportions of Specific Target Accident Types

In some circumstances it may be advantageous to examine the proportion of certain accident types out of all accidents at a site. Ranking sites by some measure of this overrepresentation may be attractive when, for example, at an intersection, the side road traffic is not known and therefore SPFs provide poor predictions. This relatively new methodology, based on the binomial test, is outlined in two recent papers by Kononov and Janson (2002) and Kononov (2002).

In this methodology, if an accident occurs, its probability to be of type *a* is P_a. (Accidents can be thought to occur as a sequence of Bernoulli trials. There exist only two outcomes for each trial, namely, if an accident of a given type has occurred or not. Furthermore, each trial is independent of all others, and the number of trials is equal to the total number of accidents.) The probability that (given the number of accidents of type *a*,) n_a accidents of type *a* occur out of a total of *N* accidents, is given by the binomial distribution:

$$P(N_a=n_a)N,P_a) = \binom{N}{n_a} P_a^{n_a}(1-P_a)^{N-n_a} \qquad \textbf{(14)}$$

where

P_a = probability that an accident is of type *a*

and

$$\binom{N}{n_a} = \frac{N!}{(N-n_a)!\,n_a!} \qquad \textbf{(15)}$$

The probability that n_a or more accidents will be observed out of *N* total accidents can be computed as:

$$P(N_a\$n_a) = 1-\sum_{0}^{n_a-1}\frac{N!}{(N-i)!\,i!}P_a^{\,i}(1-P_a)^{N-i} \qquad \textbf{(16)}$$

By setting some critical level of $P(N_a\$n_a)$, sites which are experiencing an above normal proportion of a given accident type can be identified for a detailed safety study.

This test would be applied to all accident types of interest. To make the test operational, the user will have to decide on the "critical level" (significance), but one cannot do so intelligently without an estimate of the statistical power of the approach, that is, the ability to detect true and practical deviations from a norm. Therefore, some development work will be needed. As an example, consider an urban 4-legged stop-controlled intersection with the following 5-year accident history:

N = 30 accidents in total
n_a = 15 right-angle accidents

If, for similar intersections, right-angle accidents are expected to be 35% of total accidents *(P_a = 0.35)*, we would like to know if right-angle accidents are overrepresented at this site. We would calculate:

$$P(N_a \geq 15) = 1 - \sum_{0}^{15-1} \frac{30!}{(30-i)!\,i!} 0.35^{i} (1-0.35)^{30-i} \qquad \textbf{(17)}$$

$$= 0.03$$

Thus, there is only a 3-percent chance of the intersection experiencing 15 right-angle accidents or more out of 30. One might conclude then that right-angle accidents are overrepresented at this site, and therefore a detailed safety investigation is warranted.

4.6.3 Screening of High Frequency of Specific Target Accidents

The methodology for screening of specific target accidents is the same as discussed in Section 4.5 on "Performing Basic Network Screening Calculations and Manipulations." The challenge in screening for specific target accidents is to define these accident types in available databases and to assemble or derive SPFs for these accident types. If models for a desired accident type are not available directly, a suitable SPF for total accidents can be modified by applying a multiplier calculated from the available data as the number of target accidents divided by the number of total accidents. A preliminary list of target accident types to be screened for is presented in Appendix F.

4.6.4 Identification of "Corridors With Promise"

The FHWA COTR and the MRI principal investigator participated in a safety management scanning tour as part of Task G. They found that several European countries have oriented their safety programs toward route-based improvements rather than spot improvements. These countries have concluded that most of their safety problems at spot locations have been fixed so safety improvements from roadway and traffic engineering should come primarily through route or areawide improvements. Safety management in the U.S. has not progressed to the point of eliminating the need for spot improvements. As a result, the site-specific network screening process presented in Section 4.5 of this white paper is clearly needed. However, it is recommended that a network screening process for “corridors with promise” be developed to identify extended roadway sections as potential sites for corridor- or route-based improvements.

No corridor screening process currently exists. Formal methods for the identification of corridors with promise will be explored as part of *SafetyAnalyst* research. At the very least, the identification of corridors with promise would involve plotting highly ranked spot sites on a map and performing a visual inspection for high concentrations of such sites. However, the development of a formal screening algorithm for corridors is also being considered.

4.6.5 Use of Sliding-Window Approach for Roadway Segments

A sliding-window approach will be provided for use by highway agencies that use such an approach currently and would prefer to stay with such an approach. The sliding-window approach developed for *SafetyAnalyst* will incorporate SPFs to determine expected accident frequencies and the EB approach to combining expected and observed accident frequencies.

In the sliding-window approach, a sliding window of user-specified length will move forward in increments of user-specified size along each roadway being screened. For example, the user might choose to specify a sliding window of 0.3 mi in length that moves forward in increments of 0.1 mi. This means that, in each position, the 0.3-mi segment overlaps by 0.2 mi with the previous segment and by 0.2 mi with the next segment to be screened.

The sliding-window approach will be applied to screening of roadway segment (i.e., nonintersection-related) accidents. The primary difference between the sliding-window approach and the homogeneous section approach is that the roadway within the sliding window at any given position may be nonhomogeneous. A combination of SPFs, rather than a single SPF, may need to be applied to estimate the expected accident frequency for the roadway section within the sliding window.

For any given position of the sliding window, the network screening tool will:

- determine the observed accident frequency from the *SafetyAnalyst* data files for the roadway section within the sliding window
- estimate the expected accident frequency using SPFs for the specific roadway types within the sliding window
- combine the expected and observed accident frequencies using the EB procedure
- determine the value of the user-specified screening measure (PSI based on expected accident frequency, PSI based on excess accident frequency, or prospective cost-effectiveness)
- rank sites based on the user-specified screening measure

4.7 Request Output Options and Produce Output

The primary output of the network screening tool is a ranked list of sites and the expected annual accident frequency for each location or the value of whichever other ranking criterion was used. The sites ranked would be those of the types specified by the *SafetyAnalyst* user in input data. The total number of sites ranked would be displayed to the user, and the user would be able to specify the number of highest ranked sites to be displayed. The ranking criterion would be that selected by the user at the onset of the screening analysis.

The output screen should allow the user the following capabilities:

- Print the list of ranked sites, including basic site characteristics such as:
  - Site type (intersection, road segment, freeway on-ramp etc.)
  - Environment (urban, rural)
  - Location
  - Traffic control (signalized, stop-controlled, etc.)
  - Number of lanes
  - Auxiliary lanes
  - Access control
  - Road surface type
  - Shoulder treatments
  - Traffic volumes
  - Turning restrictions

  As more variables become available in the databases input by the user, this list will expand to include them. An option will be provided for the user to customize this analysis report.

- Save the list of ranked sites in a data file, so that it can be retrieved by the user in later *SafetyAnalyst* tools; by this means, users could later select specific sites from the ranked list for consideration by the diagnosis and countermeasure selection tools
- Return to the appropriate input screen to re-rank the same sites by a different criterion
- Return to the appropriate input screen to begin a new analysis
- Eliminate specific sites from the list before further analysis is undertaken

As mentioned earlier, a potential nontraditional capability is allowing for the identification of "corridors with promise." This could be done graphically by plotting the ranked sites on a GIS display with various colors for ranges of ranking values, e.g., 1 to 10 are red, 11 to 20 are yellow, etc. Potential procedures for GIS integration will be discussed with the software development contractor when hired.

All output will be shown on the display and the user may opt to either print or save it in a file for further processing after customizing. The output format will be developed in conjunction with the other *SafetyAnalyst* tools that will use the network screening output as input data.

Section 5
Planned Development Activities

This section itemizes planned development activities, with a view to planning the next steps and scheduling the work to be done over the course of the project. The level of detail presented is consistent with this purpose.

5.1 Assembly of Background Information

The *SafetyAnalyst* network screening tool should fit well with other current safety analysis related efforts. It is important for *SafetyAnalyst* development to have current knowledge of these efforts and how *SafetyAnalyst* could and should be related. Therefore, the following tasks will be performed:

1. Determine current network screening capabilities and perceived needs of state and local agencies. The *SafetyAnalyst* network screening module should, at a minimum, reproduce the existing capabilities common to most agencies although the analytical methods used will likely be different from those currently used by highway agencies.
2. Liaise with the developers of TSIMS to understand how the interim network screening module will be used in TSIMS and how the TSIMS initiative will impact the development of the network screening module.
3. Determine how best to capitalize on the network screening research sponsored by FHWA in Colorado.
4. Gather information on average safety improvement project costs and benefits for various location types.

5.2 Technical Development

The technical development of the network screening module will be a research effort making use of appropriate databases such as Highway Safety Information Service (HSIS) data that are currently available. Aspects of the network screening module development are discussed under the following topics:

- Investigation of data requirements for planned capabilities
- Development of SPFs
- Refinement and evaluation of screening and ranking methods
- Exploration of techniques for identifying corridors with promise
- General issues to be resolved

5.2.1 Investigation of Data Requirements for Planned Capabilities

Available databases will be reviewed with the aim of determining if likely data availability would make any of the planned capabilities unfeasible. While it is almost certain that some jurisdictions may need to upgrade their data collection efforts to use *SafetyAnalyst*, the possibility that some of the planned capabilities would require a level of data collection beyond the ability of the relevant agencies will be considered. This requires close coordination with the TSIMS project.

5.2.2 Development of SPFs

Key to the development and performance of the network screening module is the ability to calibrate and supply quality SPFs for the various location types to be screened. A number of research tasks will be undertaken in this regard. The research on appropriate ranking methodologies and SPF development will be closely linked.

1. Preliminary work to assess:
    - What existing SPFs are available?
    - How transferable are they to other jurisdictions?
    - What is the best procedure for transferring them?
    - What work in SPF development is required in the *SafetyAnalyst* research?
    - What SPF related needs exist for the other applications in *SafetyAnalyst* (e.g., diagnosis and countermeasure development)?
2. Thinking, consultation, and decision making on the following issues:
    - Variable set to be used
        - As many variables as practical?
        - Only AADT?
        - Both of the above?
        - Something in between?
        - A set that is possible in most states?
    - Functional form to be used
        - Will the functional form depend on the data in each state, or is it more desirable to strive for a common functional form?
        - Will the functional form strive for commonality with accident modification factors (AMFs) in IHSDM?
    - Will the calibration use the negative binomial error structure with accident counts and AADTs for each year so as to provide yearly scale parameters, or will the benefits of the EB approach be limited by a need to use only an average of the last few years of data?

- Will there be a need to use a “per-mile” overdispersion parameter for segments?

3. Assembly of required databases for development and testing of SPFs

    - Review data needs and data sources
    - Define all site classes to be addressed in the network screening tool (e.g., roadway segment types, intersection types, ramp types)
    - Define target accident groups to be addressed in network screening (e.g., specific impact types)
    - Define variables to be addressed in network screening algorithms and related SPFs
    - Define the minimum data needs for screening

4. Assemble, develop, and/or calibrate SPFs for all site classes/target accidents to be screened in SafetyAnalyst. It is envisioned at this time that disaggregation of SPFs will be desired for a number of possible combinations of the following variables or some of these variables will be included as independent variables in the SPFs:

    - Road segments
        - Terrain (flat / rolling / mountainous)
        - Functional class (freeway, arterial, collector, local, other)
        - Divided / undivided
        - Number of lanes (2, 3, 4, > 4)
    - Intersections
        - Functional class of major, minor roads (arterial / collector / local)
        - Number of legs (3-legged / 4-legged / > 4-legged)
        - Number of lanes on approaches (2x2 / 2x4 / 4x4 / other)
        - Traffic control (signalized / all way stop / one street stopped / other)
        - Presence/absence of median on major road (divided/undivided)
    - Interchange ramps
        - Ramp function (on-ramp / off-ramp / transfer)
        - Design configuration (diamond, loop, outer connection, direct or semi-direct)
    - Severity type (PDO, NFI, and FI)
    - Environment (urban / rural)
    - Type of object struck for single-vehicle accidents
    - Manner of collision for multiple-vehicle accidents (head-on / rear-end / angle / sideswipe / other)

5. Research the specifics of how a SPF recalibration procedure should be incorporated into the SafetyAnalyst network screening algorithm

5.2.3 Refinement and Evaluation of Screening/Ranking Methods

Analysis of the merits and shortcomings of various ranking methods will be undertaken. The research will also develop enhancements as needed, following on the work being done in the Colorado project. These methods are:

1. Ranking by expected accident frequency
2. Ranking by weighted expected accident frequency
3. Ranking by excess accident frequency
4. Ranking by weighted excess accident frequency
5. Ranking by prospective cost-effectiveness, particularly for entities such as intersections that have not yet been investigated in the Colorado project
6. Screening for high proportions of specific accident types

Fundamental to this task is an investigation of the extent to which it is desirable to have an option to permit the use of a combination of ranking methods.

As part of this task, the *SafetyAnalyst* team plans to develop a pre-alpha version of the software using SAS. By implementing the network screening procedures in SAS, the *SafetyAnalyst* team will test and refine the procedures using Highway Safety Information System (HSIS) data.

5.2.4 Exploration of Techniques for Identifying Corridors With Promise

Methods for identifying corridors with promise will be explored. In particular, we will determine how the basic network screening method might be adapted to address extended route segments.

5.2.5 General Issues to be Resolved

There are a number of general issues in the development of the *SafetyAnalyst* network screening module that still need to be resolved. Some of these have been mentioned previously but are repeated here for completeness. In some cases an assumption has been made on a likely resolution to provide the required information on how the network screening module would work.

1. What is the base level of information required to use *SafetyAnalyst*?
2. Who is the target user for the interim network screening module? For the final module? In particular, assuming that state agencies are the primary users, to

what extent should the network screening module be applicable to local agencies?

3. What target accident types should be addressed by the network screening module? Current highway agency systems consider various accident types including total accidents, accidents for specific severity levels, wet-pavement accidents, and alcohol-related accidents.
4. Do the planned network screening capabilities meet the expectations of potential users?
5. Is it more efficient to screen based on expected accident frequency or the expected accident frequency above a "norm" (the approach referred to earlier in this work plan as excess accident frequency)?
6. Should there be a capability for simultaneously ranking different types of locations? Is it reasonable to combine sites of different types (e.g., urban and rural sites) in network screening without accounting for the potential differences in costs and benefits for projects at the different types of sites? If not, how can the screening method accommodate these potential differences?
7. Is a map or geographical display of the network screening results necessary for *SafetyAnalyst* users?
8. What mapping or GIS interface should be provided to visually display the network screening results?
9. How can a different map be incorporated in *SafetyAnalyst* for each state/local agency that uses the *SafetyAnalyst* software? The *SafetyAnalyst* mapping functionality could possibly be developed in a generic fashion to allow the agency to provide a GIS map and "import" it into the *SafetyAnalyst* system. Regardless of the type of database used, a third-party product such as ArcView or MapInfo would be needed to actually display the map and plot the various roadway location objects.
10. How should an SPF calibration procedure be incorporated into the *SafetyAnalyst* network screening algorithm? Can a procedure analogous to that developed for the IHSDM Crash Prediction Module be used or is a different approach needed?
11. How can the network screening module be tailored to look for opportunities for safety improvements in capital projects planned for reasons other than safety?

Item 6 in the above list represents a key issue that needs to be considered. Past network screening algorithms have been quite generic in nature and have often addressed particular issues in isolation (e.g., accidents on freeway segments or wet-pavement accidents). Existing network screening approaches are suitable for ranking roadway segments or intersections of a specific type in terms of accident frequency or excess accident frequency. However, it is difficult to combine different types of sites, such as urban and rural sites, based on these criteria because urban and rural sites may differ substantially in the potential costs or potential benefits of projects. For example, urban sites generally have higher accident frequencies than rural sites, so they would generally rank higher on any combined network screening list based on accident frequency;

however, if safety improvement projects in urban areas are generally more expensive to construct than safety improvement projects in rural areas, it could be potentially misleading to fail to account for this. The Colorado DOT project has proposed to address this issue by ranking sites based on prospective cost-effectiveness. There are two ways in which this could be done:

- Obtain highway agency data on the average cost of safety improvements for different highway types and use those averages as weights in ranking lists of sites of varying types.
- Estimate both the project costs and prospective benefits for specific sites based on a tentative diagnosis made by the network screening algorithm from available accident data (without the benefit of the office and field investigations that will be undertaken in using the *SafetyAnalyst* diagnosis tool) and use these estimates in assigning a prospective cost-effectiveness for ranking specific sites.

As discussed above, the efficiency of this approach needs to be further investigated. Combined priority lists, including different site types, are potentially important because the diagnostic studies that are essential to identifying the improvements actually needed at specific sites are expensive as can be the consequences of ignoring sites in dire need of safety treatment. Since the number of studies that can be undertaken is limited, it is important to target efforts with the greatest possible efficiency.

Some of the development activities described above are being addressed in the ongoing Colorado DOT project and in other research efforts. Where applicable, we will (with permission from the sponsors of those other efforts) use the results of those efforts in developing the *SafetyAnalyst* network screening module. Thus, we will seek, where possible, to avoid reinventing the wheel and to minimize the development effort required for *SafetyAnalyst*. However, a critical review and, where possible, an empirical validation of all external results will be conducted before incorporating them in *SafetyAnalyst* to assure that all aspects of *SafetyAnalyst* are well supported and widely applicable.

5.2.6 Summary of Gaps to Be Filled

This subsection summarizes, in point form, the gaps to be filled in the technical developmental work of the network screening module:

1. Completing the suite of SPFs. This will require review for suitability of available SPFs and the calibration of new ones. In the process, examination of unresolved issues such as the variation of the overdispersion parameter across jurisdictions and other factors such as section length and ADT will be investigated.
2. Further testing and enhancement of screening methods being developed in the Colorado project. This includes:

- Evaluation of screening for high proportions of specific target accident types
- Evaluation of ranking by prospective cost-effectiveness
- Investigation of the definition of potential for safety improvement
- Investigation of the definition of what is a site, in reference to road segments

3. Development of a pre-alpha version of the software using SAS
4. Applicability to other jurisdictions and location types of the methods being developed in the Colorado project for road segments
5. Development of a GIS interface to visually display the network screening results
6. Development of a corridor screening methodology
7. Refinement of techniques for recalibrating SPFs to apply in different time periods and jurisdictions.

Section 6
References

Harwood, D. W., F. M. Council, E. Hauer, W. E. Hughes, and A. Vogt. *Prediction of the Expected Safety Performance of Two-Lane Highways*, Report No. FHWA-RD-99-207, Federal Highway Administration, December 2000.

Harwood, D. W., K. M. Bauer, I. B. Potts, D. J. Torbic, K. R. Richard, E. R. Kohlman Rabbani, E. Hauer, and L. Elefteriadon, *Safety Effectiveness of Intersection Left- and Right-turn Lanes*, Report No. FHWA-RD-02-089, Federal Highway Administration, July 2002.

Hauer, E. Working Paper 10. "Screening: Methods and Software," Draft report for FHWA and Colorado DOT, February 28, 2000.

Hauer, E. *Observational Before-After Studies in Road Safety*, Pergamon Press, 1997

Hauer, E. *Detection of Safety Deterioration in a Series of Accident Counts*. Transportation Research Record 1542, 38-43, 1996a.

Hauer, E. *Statistical Test of the Difference Between Expected Accident Frequencies.* Transportation Research Record 1542, 24-29, 1996b.

Hauer, E., D. Harwood, F. Council, and M. S. Griffith. "Estimating Safety by the Empirical Bayes Method: A Tutorial." Transportation Research Record, in press (presented at TRB annual meeting in January 2002a).

Hauer, E., J. Kononov, B. Allery, and M. Griffith. "Screening the Road Network for Sites with Promise," Transportation Research Record, in press (presented at the TRB annual meeting, January 2002b).

Kononov, J. "Use of Direct Diagnostics and Pattern Recognition Methodologies in Identifying Locations with Potential for Accident Reductions," Transportation Research Record, in press (presented at the TRB annual meeting, January 2002).

Kononov, J. and R. Janson. "Diagnostic Methodology for the Detection of Safety Problems at Intersections," Transportation Research Record, in press (presented at the TRB annual meeting, January 2002).

Pendleton, O. *Evaluation of accident analysis methodology,* Report No. FHWA-RD-96-039, Federal Highway Administration, 1996.

Persaud, B.N. *Statistical Methods in Highway Safety Analysis,* National Cooperative Highway Research Program Synthesis 294, Transportation Research Board, 2001.

Appendix A

A Primer on the Pitfalls of Screening Based on Observed Accident Counts and/or Rates

(Excerpt from Persaud B.N. "Statistical Methods in Highway Safety Analysis." National Cooperative Highway Research Program Synthesis 294, Transportation Research Board, 2001)

Regression-to-the-Mean and the Problem With Using the Accident Count

A common method for screening is a simple ranking on the basis of the observed accident counts. It is the simplest of techniques and requires much less data than the more sophisticated techniques but suffers from the regression-to-the-mean bias in which an unusually high count is likely to decrease subsequently even if no improvement were implemented. Therefore, a site with such counts may not be in need of improvement. Conversely, a truly hazardous site may have a randomly low observed count and incorrectly escape detection as a result.

The difficulties caused by the regression-to-the-mean phenomenon have been well documented in published literature and are now recognized by many safety analysts, but for the uninitiated an illustrative example may still be in order here. Table A1 is assembled from data presented in Hauer (1997) for 1072 San Francisco intersections with specific numbers of accidents in 1974 to 76. For the same intersections in each row, the average number of accidents per intersection for 1977 is also shown. There was no real change in safety at these intersections between 1974 to 76 and 1977 in that accidents averaged over all intersections remained essentially constant over the years at about 1.1 accidents per intersection per year. Yet, intersections that had 4 or more accidents in 1974 to 76 (more than the average of 1.1 per year) recorded substantial reductions in accidents the following year, indicating that at least some of them are not really hazardous, or at the very least that using the count of 4 or more accidents in 3 years to identify sites for treatment is not a very efficient procedure. Conversely, those with 3 or less (i.e., less than the average of 1.1 per year), which are likely to be deemed safe in an accident count screening procedure, experienced an increase, indicating that the low count in 1974 to 76 was an anomaly and that some of these sites may actually be worthy of investigation.

Misplaced faith in the accident count based screening method can be perpetuated by the conduct of a simple before-after comparison of sites treated, and ignoring the regression-to-the-mean effects. For example, if the 54 intersections with 6 accidents in 3 years (2.0 per year) were treated at the end of the 1974 to 76 period and recorded, say, 1.12 accidents per intersection in 1977, a conventional before-after comparison would estimate the treatment effect as a reduction of 44% (=100(2-1.12)/2). Yet, as the last column in Table A-1 shows, this would be a gross overestimate since half of that apparent reduction would have been due to regression-to-the-mean.

Table A-1. Illustrating the Regression-to-the-Mean Phenomenon

Number of intersections	Accidents/ intersection in 1974-76	Accidents/year/ intersection in 1974-76	Accidents/ intersection in 1977	% Change
256	0	0	0.25	Large increase
218	1	0.33	0.55	67%
173	2	0.67	0.70	Small increase
121	3	1.00	1.04	Small increase
97	4	1.33	1.08	–19%
70	5	1.67	1.33	–20%
54	6	2.00	1.56	–22%
32	7	2.33	2.25	–3%
29	8	2.67	1.62	–39%

The Problem With Accident Rates

In some jurisdictions, accident rate is used directly or indirectly as a hazard measure to flag locations for safety investigation. AADTs are used directly in the computation of this measure, i.e., accident rate = accident frequency/AADT (or some scalar multiple of this).

If accident rates are based on the observed counts, then the regression-to-the-mean difficulty discussed above will still apply. In addition, there is an additional problem that renders this method of screening dubious. The problem, as the extensive literature on SPFs shows, is that the relationship between accident frequency and AADT is not linear (Pendleton, 1996). Figure A-1, which depicts the SPF for injury accidents for two-lane rural roads in Ontario illustrates the inherent nonlinearity and the difficulties with the linearity assumption. The relationship depicted is of the form:

$$Accidents/km/unit\ of\ time = a(AADT)^b$$

where:

a = 0.00398
b = 0.812

are regression coefficients calibrated from data.

A value of b = 1 would have indicated a linear relationship.

The nonlinearity depicted in Figure A-1 points to an inherent flaw in the use of accident rate as a measure of safety. Specifically, comparing accident rates of two entities at different traffic levels to judge relative safety may lead to erroneous conclusions. According to Figure A-1, the accident rate (the slope of a line from the origin to a point on the curve) is expected to be lower at higher traffic volumes. Thus, saying that when two rates are equal they indicate equivalent levels of hazard may be completely false if different AADT levels are involved. The upshot of all this is that the use of accident rate to compare sites in regard to their safety levels is potentially problematic. When the slope of the accidents/AADT relationship is decreasing with increasing traffic volume levels, as is often the case, screening by accident rates will tend to identify low AADT sites for further investigation. The most valid basis of comparison using accident rates is for the relatively rare cases when the traffic volume levels are the same or when the relationship between accidents and AADT is linear.

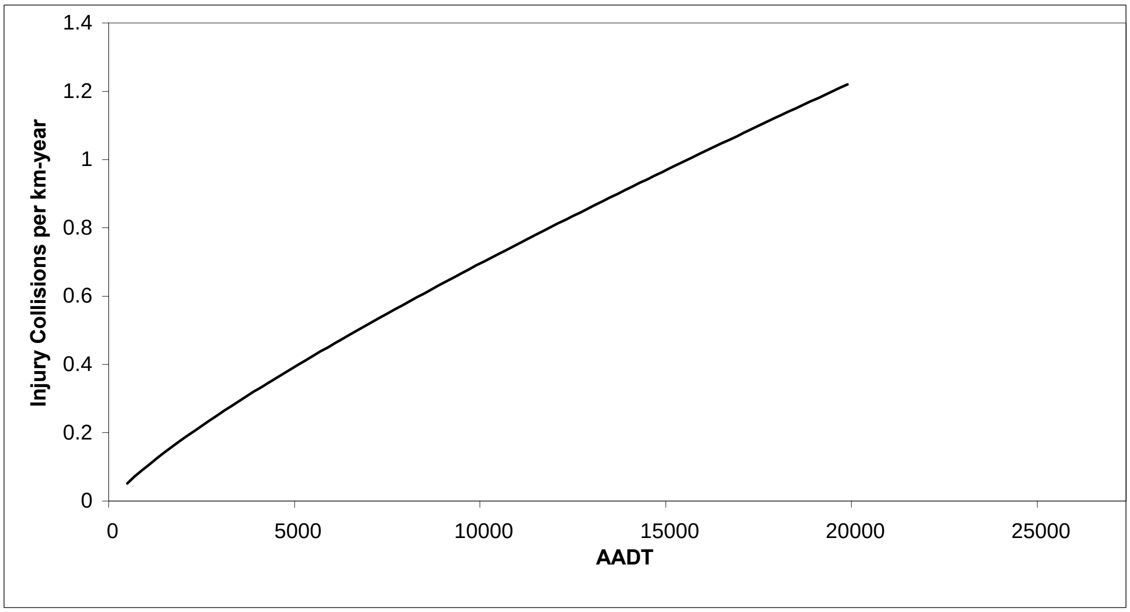

Figure A-1. Safety Performance Function for Two-Lane Rural Arterial Highways in Ontario

Appendix B

Preliminary List of Desired Data Requirements

Preliminary List of Desired Data Requirements

Traffic Accident Data

Accident location (milepost, link/node/offset, segment/offset, GPS coordinates)
Date (day/month/year)
Day of week
Time (at least to the nearest hour)
Accident severity (F/I/PDO or K/A/B/C/PDO)
Tow-away indicator (yes/no)
Relationship to junction (at intersection/intersection-related/driveway-related/at railroad-highway grade crossing/grade-crossing-related/ramp/speed-change lane/not related to junction)
Light condition
Weather
Pavement surface condition
Number of vehicles involved
Accident type (based on type object struck for single-vehicle accidents and manner of collision for multiple-vehicle accidents)
Initial direction of travel (for at least Vehicles 1 and 2)
Intended maneuver (for at least Vehicles 1 and 2)
Driver age (for at least Vehicles 1 and 2)
Vehicle types involved (for all involved vehicles)
First harmful event
Most harmful event
Object struck
Ran-off road indicator (yes/no)
Pedestrian indicator (yes/no)
Bicycle indicator (yes/no)

Roadway Segment Inventory Data

Segment location (beginning and ending points; linkable to accident data)
Segment length
Area type (rural/urban)
Basic number of lanes
Auxiliary lanes (TWLTL/passing lane/climbing lane/other auxiliary lane)
Median type
Median width
Lane width
Shoulder type
Shoulder width
Access point density
Average daily traffic (ADT)
Peak hour volume/design hour volume
Percent heavy vehicles
Speed (85th percentile or posted speed)

Intersection Inventory Data

Intersection location (linkable to roadway segment and accident data)
Area type (rural/urban)
Number of legs
Type of traffic control (signalized/two-way STOP/all-way STOP/two-way YIELD/uncontrolled)
Number of through lanes on major road (includes shared lanes)
Median type on major road
Median width on major road
Left-turn lanes on major road
Right-turn lanes on major road
Number of through lanes on minor road (included shared lanes)
Median type on minor road
Median width on minor road
Left-turn lanes on minor road
Right-turn lanes on minor road
Traffic volume (ADT) on major road
Peak hour volume/design hour volume on major road
Traffic volume (ADT) on minor road
Peak hour volume/design hour volume on minor road
Turning volumes

Interchange Ramp Inventory Data

Ramp location (linkable to mainline roadway segment and accident data)
Area type (rural/urban)
Ramp length
Ramp type (diamond/loop/outer connection/directional/semidirectional)
Type of connection at either end (mainline acceleration lane/mainline deceleration lane/mainline weaving area/C-D road/other ramp)
Speed (85th percentile, posted, or advisory speed)

Other Data Files

Horizontal curve data
Grade and vertical curve data
Railroad-highway grade crossing inventory data

Cost Data

Average cost of safety improvement projects where available disaggregated as far as possible
Average cost of accidents of different severities

SPF Data

Functional form
Parameters (regression coefficients and overdispersion parameter)
Valid parameter ranges

Appendix C

Guidance for Developing User-Provided SPFs and Instructions for Recalibrating Default SPFs

A major component of the input elements discussed in Section 4 is a database of SPFs. *SafetyAnalyst* will be provided with a set of default SPFs that will be periodically updated by FHWA, at least with each new revision of the software. In those cases where an agency chooses to use its own SPFs or plans to recalibrate default SPFs to fit its needs, a number of steps need to be taken to ensure that user-provided SPFs were obtained in a correct fashion.

This appendix provides guidance for developing SPFs when default SPFs are not used and instructions for recalibrating default SPFs to suit an agency's particular situation.

C.1 Guidance for Developing User-Supplied SPFs

From existing databases of accidents, traffic volumes, and roadway geometries, an agency may develop their own SPFs using commercially available software. For an analyst with some experience or under the supervision of a statistician or safety researcher, this approach will be quite straightforward when using simple and common functional forms in which the SPF is of a multiplicative kind, each variable having its functional form and these multiplying each other. For roadway segments, examples of these common forms are:

$$\text{Accidents/year} = (\text{segment length})\ \acute{a}(x_1)\hat{a}_1(x_2)\hat{a}_2\ldots\ (x_n)\hat{a}_n \qquad \textbf{(1)}$$

$$\text{Accidents/year} = (\text{segment length})\ \exp(\acute{a}+\ \hat{a}_1x_1+\ \hat{a}_2x_2+\ldots..\ \hat{a}_nx_n) \qquad \textbf{(2)}$$

where:

$x_1 ,\ldots,x_n$	=	traffic and geometric variables such as average annual daily traffic (AADT) and lane width
$\acute{a}$ and $\hat{a}_1,\ldots,\hat{a}_n$	=	coefficients estimated in the model calibration procedure

However, future SPFs may be more complex than suggested by Equations (1) and (2). For example, Hauer (2002) has proposed a revised, unconventional approach to developing SPFs for two-lane road segments for application in the Interactive Highway Safety Design Model (IHSDM); no decision has yet been made about whether to adopt this approach in IHSDM.

It is possible that in some cases the influence of a variable will be represented by a few regression parameters, not a function. For example, variations of the above equations may have several parameters α_{year}, one for every year used in the screening. Segment length also may have to be represented by a more complex function, and not just as a multiplier. Thus, the process of developing user-supplied SPFs may be quite complex. Because of this complexity, verification is essential. One verification level is to show the function for each variable introduced as a graph, first before the parameters are entered in some generic form and then after the parameters are entered in their true form. Another verification level would be to make validation an integral part of the process of entering the SPF. The premise is that if the user has an SPF, it must be based on the user's data, the data that also serve for screening. Therefore, the user's SPF must fit the data and have a ratio of observed to predicted accidents of close to 1.

In addition to the functional form and the coefficients of the regression equation, the user will also have to specify the value of *k*, an overdispersion parameter estimated during model calibration. The regression coefficients and the overdispersion parameter are essential in the network screening calculations (Hauer et al., 2002a).

C.2 Instructions for Recalibrating Base SPFs

The procedure for recalibrating default SPFs (provided by *SafetyAnalyst*) to a particular situation is *essentially* as per the following illustration. Consider an SPF for total accidents being recalibrated for urban 4-legged intersections. Suppose the default SPF for total accidents is:

$$\textbf{Total accidents/year} = \mathbf{0.00005(AADT_1)^{0.750}(AADT_2)^{0.350}} \qquad (3)$$

where:

$AADT_1$ = major road AADT
$AADT_2$ = minor road AADT

Assume the data to be screened consist of 200 urban 4-legged intersections with the following accident history (i.e., observed accident frequencies):

Year 1: 150 total accidents
Year 2: 130 total accidents
Year 3: 165 total accidents

Step 1: Apply the SPF from Equation (3) to estimate the number of accidents, separately for years 1 to 3, at each of the 200 intersections. Use the AADTs for the respective year.

Step 2: For each year, calculate a yearly calibration factor, C_i, by dividing the sum over all sites of the observed number of accidents in that year by the sum of the predicted number of accidents in that year:

$$C_i = \frac{\sum Observed\ accidents_i}{\sum Predicted\ accidents_i} \qquad (4)$$

In this case, suppose the sums for all sites of the yearly predictions were:

Year 1: 134.50 total accidents
Year 2: 140.75 total accidents

Year 3: 150.55 total accidents

It follows that:

$C_1 = 150/134.50 = 1.12$
$C_2 = 130/140.75 = 0.92$
$C_3 = 165/150.55 = 1.10$

Step 3: Add the calibration factors to the SPF as a multiplier for each year. The recalibrated SPF is then:

$$\text{Accidents/year} = (C_i)(0.00005)(ADT_1)^{0.750}(ADT_2)^{0.350} \quad \textbf{(5)}$$

that is, three SPFs in fact were developed in this particular case.

Step 4: Using the recalibrated SPF for each year [Equation (5)], estimate the predicted number of accidents, P, for each site and each year. Steps 1 through 5 are summarized in Table C-1.

Step 5: Recalibrate the overdispersion parameter, *k*. *(It is possible that, as a result of the research, this step may not be necessary. It is documented here to provide a feel for what is involved should it be necessary.)*

1. For each site, calculate the total number of observed accidents, *0*, across all three years.
2. Similarly, for each site, calculate the total number of predicted accidents, *P*, across all three years. Also compute P^2 for each site.
3. For each site, determine the value of the squared residual (*SR*):

$$SR = (P - O)^2 \quad \textbf{(6)}$$

4. Subtract the value of P from the squared residual (SR). This gives an estimate of P^2/k:

$$[\text{Estimate of } P^2/k\ ] = SR - P \quad \textbf{(7)}$$

5. Fit a straight line to the data with P^2/k as the dependent variable and P^2 as the independent variable, forcing the line through the origin. Thus, in this example, a straight line, forced through the origin, will be fit to 200 pairs of [P^2, *(SR-P)*] data points. An ordinary least squared regression procedure such as that provided by Excel should suffice.
6. The inverse of the slope of the fitted regression line is an estimate of *k*.

Table C-2 summarizes Step 5 calculations. The highlighted fourth and sixth columns show the data used for estimating the slope of the straight line to obtain an estimate of the overdispersion parameter.

Table C-1. Example Calculations of Yearly Calibration Factors and Final Predicted Accident Frequencies

Intersection No.	Observed accident frequency			Predicted accident frequency using default SPF (Eq. 3)			Predicted accident frequency using recalibrated SPF (Eq. 5)		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
1	$O_{1,1}$	$O_{1,2}$	$O_{1,3}$				$P_{1,1}$	$P_{1,2}$	$P_{1,3}$
2	$O_{2,1}$	$O_{2,2}$	$O_{2,3}$				$P_{2,1}$	$P_{2,2}$	$P_{2,3}$
3									
.									
.									
i									
.									
.									
200	$O_{200,1}$	$O_{200,2}$	$O_{200,3}$				$P_{200,1}$	$P_{200,2}$	$P_{200,3}$
Total	150	130	165	134.50	140.75	150.55			

Table C-2. Example Calculations for Recalibrated Overdispersion Parameter

Intersection No.	Accident frequencies—total over 3 years			Squared residual $(P-O)^2$	Estimate of P^2/k
	Observed	Predicted	Predicted^2		
1	O_1	P_1	$\mathbf{P_1^2}$	SR_1	$\mathbf{SR_1 - P_1}$
2	O_2	P_2	$\mathbf{P_2^2}$	SR_2	$\mathbf{SR_2 - P_2}$
3					
.					
.					
i					
.					
.					
200	O_{200}	P_{200}	$\mathbf{P_{200}^2}$	SR_{200}	$\mathbf{SR_{200} - P_{200}}$

Appendix D

Example of PSI Calculations Based on a Site's Expected Accident Frequency

Example of PSI Calculations on a Site's Expected Accident Frequency

For this example, adapted from Hauer (2000), consider a 1.73-mi long rural two-lane road section in rolling terrain with a 13-year history (1989-2001) of AADTs and accidents. The road section thus consists of 18 subsections: 17 of 0.10 mi and 1 of 0.03 mi. The model equation is "AADT β_0 exp(β_1 AADT). The model parameters α, β_0, β_1, and the overdispersion parameter, *k*, for Total Accidents and for Injury Accidents are shown in Table D-1. AADTs for each of the 13 years are shown as well. PDO, NFI, and FI accident frequencies are shown in Table D-2.

Table D-1. Model Parameters for Total Accidents and Injury Accidents in Rolling Terrain

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
AADT	4,650	4,700	4,750	4,800	4,700	4,600	4,500	4,400	4,400	4,600	4,800	4,900	5,000
α, Total Accidents	2.172	2.367	2.129	1.873	1.888	1.875	1.656	1.763	1.795	1.849	1.905	2.183	1.937
α, Injury Accidents	0.876	0.871	0.833	0.792	0.851	0.774	0.687	0.710	0.736	0.690	0.701	0.718	0.653

	k	$\$_0$	$\$_1$
Total Accidents	0.208/mi or 0.294	0.7112	0.5321
Injury Accidents	0.190/mi or 0.246	0.6834	0.6277

Table D-2. Yearly Accident Frequencies for Each Subsection

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	PDO Accident Frequencies																	
1989	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
1991	0	0	0	0	0	0	0	0	0	1	1	0	0	0	2	0	0	0
1992	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
1993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	0	0	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0
1996	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
1997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1998	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0
1999	0	0	0	0	0	1	0	0	1	2	0	0	1	0	1	1	0	0
2000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Year	NFI Accident Frequencies																	
1989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1992	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1
1993	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1994	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
1995	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1996	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1998	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1999	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2000	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Year	FI Accident Frequencies																	
1989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1991	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1992	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1998	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The calculations for Steps 1 through 6 shown in Section 4.5.1 are now implemented. The results for Step 1 through 3 are shown in Table D-3.

Table D-3. Results From Example Calculations

		Estimated Models				From estimated models					
		Total		Injury (NFI&FI)		PDO	NFI	FI	PDO	NFI	FI
Y	X	"$_y$	$E(K_y)$	"$_y$	$E(K_y)$		$E(K_y)$			Cy	
(Years)	(ADT/10,000)		Acc./m-y		Acc./m-y		1-R=0.944	R=0.056			
1989	0.465	2.172	1.614	0.876	0.695	0.919	0.656	0.039	1.000	1.000	1.000
1990	0.470	2.367	1.777	0.871	0.698	1.078	0.660	0.039	1.174	1.005	1.005
1991	0.475	2.129	1.614	0.833	0.675	0.940	0.637	0.037	1.023	0.971	0.971
1992	0.480	1.873	1.435	0.792	0.648	0.786	0.612	0.036	0.856	0.933	0.933
1993	0.470	1.888	1.417	0.851	0.682	0.735	0.644	0.038	0.800	0.982	0.982
1994	0.460	1.875	1.379	0.774	0.608	0.771	0.574	0.034	0.839	0.874	0.874
1995	0.450	1.656	1.192	0.687	0.528	0.664	0.499	0.029	0.723	0.760	0.760
1996	0.440	1.763	1.243	0.710	0.534	0.709	0.504	0.030	0.771	0.768	0.768
1997	0.440	1.795	1.265	0.736	0.554	0.712	0.523	0.031	0.775	0.796	0.796
1998	0.460	1.849	1.360	0.690	0.542	0.818	0.512	0.030	0.890	0.779	0.779
1999	0.480	1.905	1.459	0.701	0.574	0.885	0.542	0.032	0.964	0.826	0.826
2000	0.490	2.183	1.706	0.718	0.600	1.106	0.566	0.033	1.204	0.863	0.863
2001	0.500	1.937	1.544	0.653	0.557	0.987	0.526	0.031	1.075	0.801	0.801
Sums			19.004		7.894	11.110	7.455	0.439	12.094	11.357	11.357
$\$_0$		0.7112		0.6834							
$\$_1$		0.5321		0.6277				*w* =	0.322	0.404	0.978
k (per mi)		0.208		0.190				Note: *k* for "injury" used for all three *w*			

Note that the $E(K_y)$ in Table D-3 are for one mile. Since k is also "per mile," the weights w can be computed from these values directly. The computations of Steps 4 through 6 for PDO, NFI, and FI accidents are shown in Tables D-4, D-5, and D-6, respectively.

Of interest are several accident profiles. These profiles are shown in Table D-7. The computations are explained next.

- Let S_K denote the sum of $\hat{K}_Y$ for PDO, NFI, and FI accidents. Thus,

$$S_K = \hat{K}_{Y(PDO)} + \hat{K}_{Y(NFI)} + \hat{K}_{Y(FI)} \qquad \textbf{(1)}$$

- Let $VAR(S_K)$ denote the sum of $VAR(\hat{K}_Y)$ for PDO, NFI, and FI accidents. Thus,

$$VAR(S_K) = VAR(\hat{K}_{Y(PDO)}) + VAR(\hat{K}_{Y(NFI)}) + VAR(\hat{K}_{Y(FI)}) \qquad \textbf{(2)}$$

- Let RCS_K denote the "cost-weighted" sum of $\hat{K}_Y$, where the relative cost values RC_{PDO}, RC_{NFI}, and RC_{FI} serve as weights. The relative cost values used in this example for PDO, NFI, and FI accidents are 1, 4, and 200. Thus,

$$RCS_K = RC_{PDO}\hat{K}_{Y(PDO)} + RC_{NFI}\hat{K}_{Y(NFI)} + RC_{FI}\hat{K}_{Y(FI)} \qquad \textbf{(3)}$$

Table D-4. Estimate of Expected PDO Accidents in Last Year (2001) and Variance of Estimates

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	PDO Accident Frequencies																	
1989	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
–	–	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–
2000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Sum	2	1	0	2	2	1	2	1	2	3	3	2	2	0	3	2	0	0
$\hat{K}_1$	0.142	0.086	0.030	0.142	0.142	0.086	0.142	0.086	0.142	0.198	0.198	0.142	0.142	0.030	0.198	0.142	0.030	0.009
$\hat{K}_Y$	0.152	0.092	0.032	0.152	0.152	0.092	0.152	0.092	0.152	0.213	0.213	0.152	0.152	0.032	0.213	0.152	0.032	0.010
$VAR(\hat{K}_Y)$	0.0092	0.0055	0.0019	0.0092	0.0092	0.0055	0.0092	0.0055	0.0092	0.0128	0.0128	0.0092	0.0092	0.0019	0.0128	0.0092	0.0019	0.0006

Table D-5. Estimate of Expected NFI Accidents in Last Year (2001) and Variance of Estimates

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	NFI Accident Frequencies																	
1989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
–	–	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–
2000	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Sum	1	0	0	2	1	1	1	0	1	2	1	2	3	0	0	0	1	1
$\hat{K}_1$	0.079	0.027	0.027	0.131	0.079	0.079	0.079	0.027	0.079	0.131	0.079	0.131	0.184	0.027	0.027	0.027	0.079	0.060
$\hat{K}_Y$	0.063	0.021	0.021	0.105	0.063	0.063	0.063	0.021	0.063	0.105	0.063	0.105	0.147	0.021	0.021	0.021	0.063	0.048
$VAR(\hat{K}_Y)$	0.0027	0.0009	0.0009	0.0044	0.0027	0.0027	0.0027	0.0009	0.0027	0.0044	0.0027	0.0044	0.0062	0.0009	0.0009	0.0009	0.0027	0.0020

Table D-6. Estimate of Expected FI Accidents in Last Year (2001) and Variance of Estimates

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	FI Accident Frequencies																	
1989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
$\hat{K}_1$	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.004	0.001
$\hat{K}_Y$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.005	0.003	0.003	0.003	0.003	0.001
$VAR(\hat{K}_Y)$	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	7.10E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	1.41E-06

- Let $VAR(RCS_K)$ denote the "cost-weighted" sum of $VAR(\hat{K}_Y)$ for PDO, NFI, and FI accidents. Thus,

$$VAR(RCS_K) = RC_{PDO}^2 VAR(\hat{K}_{Y(PDO)}) + RC_{NFI}^2 VAR(\hat{K}_{Y(NFI)}) + RC_{FI}^2 VAR(\hat{K}_{Y(FI)}) \qquad \textbf{(4)}$$

Table D-7. Summed Unweighted and Cost-Weighted Accident Frequency Profiles

Subsection	S_K for 2001	$VAR\ (S_K)$ for 2001	RCS_K for 2001	$VAR\ (RCS_K)$ for 2001
1	0.219	0.0118	1.010	0.2393
2	0.116	0.0064	0.782	0.2074
3	0.056	0.0028	0.721	0.2038
4	0.261	0.0136	1.178	0.2676
5	0.219	0.0118	1.010	0.2393
6	0.156	0.0082	0.950	0.2357
7	0.219	0.0118	1.010	0.2393
8	0.116	0.0064	0.782	0.2074
9	0.219	0.0118	1.010	0.2393
10	0.321	0.0172	1.238	0.2712
11	0.279	0.0155	1.070	0.2429
12	0.261	0.0136	1.178	0.2676
13	0.304	0.0154	1.656	0.3920
14	0.056	0.0028	0.721	0.2038
15	0.237	0.0137	0.902	0.2147
16	0.177	0.0101	0.842	0.2111
17	0.098	0.0046	0.890	0.2321
18	0.059	0.0026	0.385	0.0894

Appendix E

Example of PSI Calculations Based on a Site's Expected Excess Accident Frequency

Example of PSI Calculations Based on a Site's Expected Excess Accident Frequency

For the example calculations shown in Appendix D, remedial projects result in an accident cost reduction that is proportional to the expected cost-weighted accident frequency. An alternative assumption is that remedial projects result in an accident cost reduction proportional to the expected *excess* of the cost-weighted accident frequency. The creation of the corresponding profile will be examined here.

We are considering the same 1.73-mi long rural two-lane road section in rolling terrain with a 13-year history (1989-2001) of AADTs and accidents as used in Appendix D. The same model equations and data already given there apply here. Similarly, the results of computations for Steps 1 through 3 shown in Section 4.5.1 carry over without change. The computations of Steps 4 through 6 for PDO, NFI, and FI accidents are shown in Tables E-1, E-2, and E-3, respectively, where:

$$Excess = \hat{K}_Y - E(K_Y) \quad (1)$$

$$VAR[E(K_Y)] = kE(K_Y)^2 \quad (2)$$

Of interest are several excess accident profiles. These profiles are shown in Table E-4. The computations are explained next.

- Let ES_K denote the sum of excess PDO, NFI, and FI accidents. Thus,

$$ES_K = Excess_{PDO} + Excess_{NFI} + Excess_{FI} \quad (3)$$

- Let $VAR(ES_K)$ denote the sum of the $VAR(\hat{K}_Y)$ and $VAR[E(K_Y)]$ for PDO, NFI, and FI accidents. Then,

$$VAR(ES_K) = [VAR(\hat{K}_{Y(PDO)}) + VAR(E(K_{Y(PDO)})] + [VAR(\hat{K}_{Y(NFI)}) + VAR(E(K_{Y(NFI)})] + [VAR(\hat{K}_{Y(FI)}) + VAR(E(K_{Y(FI)})] \quad (4)$$

- Let $ERCS_K$ denote the "cost-weighted" sum of excess accidents where the relative cost values RC_{PDO}, RC_{NFI}, and RC_{FI} serve as weights. The relative cost values used in this example for PDO, NFI, and FI accidents are 1, 4, and 200. Thus,

$$ERCS_K = RC_{PDO}\, Excess_{PDO} + RC_{NFI}\, Excess_{NFI} + RC_{FI}\, Excess_{FI} \quad (5)$$

- Let $VAR(ERCS_K)$ denote the sum of the "cost-weighted" $VAR(\hat{K}_Y)$ and $VAR[E(K_Y)]$ for PDO, NFI, and FI accidents. Then,

$$VAR(ERCS_K) = RC_{PDO}^2 [VAR(\hat{K}_{Y(PDO)}) + VAR(E(K_{Y(PDO)})] + [RC_{NFI}^2\, VAR(\hat{K}_{Y(NFI)}) + VAR(E(K_{Y(NFI)})] + [RC_{FI}^2\, VAR(\hat{K}_{Y(FI)}) + VAR(E(K_{Y(FI)})] \quad (6)$$

Table E-1. Estimate of Expected Excess PDO Accidents in Last Year (2001) and Its Variance

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	PDO Accident Frequencies																	
1989	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
–	–	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–
2000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Sum	2	1	0	2	2	1	2	1	2	3	3	2	2	0	3	2	0	0
$\hat{K}_1$	0.142	0.086	0.030	0.142	0.142	0.086	0.142	0.086	0.142	0.198	0.198	0.142	0.142	0.030	0.198	0.142	0.030	0.009
$\hat{K}_Y$	0.152	0.092	0.032	0.152	0.152	0.092	0.152	0.092	0.152	0.213	0.213	0.152	0.152	0.032	0.213	0.152	0.032	0.010
$E(K_y)$	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.099	0.030
$Excess$	0.054	–0.007	–0.067	0.054	0.054	–0.007	0.054	–0.007	0.054	0.114	0.114	0.054	0.054	–0.067	0.114	0.054	–0.067	–0.020
$VAR(\hat{K}_Y)$	0.0092	0.0055	0.0019	0.0092	0.0092	0.0055	0.0092	0.0055	0.0092	0.0128	0.0128	0.0092	0.0092	0.0019	0.0128	0.0092	0.0019	0.006
$VAR[\mathrm{E}(K_y)]$	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0185	0.0055

Table E-2. Estimate of Expected Excess NFI Accidents in Last Year (2001) and Its Variance

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	NFI Accident Frequencies																	
1989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
–	–	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–
2000	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
Sum	1	0	0	2	1	1	1	0	1	2	1	2	3	0	0	0	1	1
$\hat{K}_1$	0.079	0.027	0.027	0.131	0.079	0.079	0.079	0.027	0.079	0.131	0.079	0.131	0.184	0.027	0.027	0.027	0.079	0.060
$\hat{K}_Y$	0.063	0.021	0.021	0.105	0.063	0.063	0.063	0.021	0.063	0.105	0.063	0.105	0.147	0.021	0.021	0.021	0.063	0.048
$E(Ky)$	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.016
Excess	0.011	-0.031	-0.031	0.053	0.011	0.011	0.011	-0.031	0.011	0.053	0.011	0.053	0.095	-0.031	-0.031	-0.031	0.011	0.033
$VAR(\hat{K}_Y)$	0.0027	0.0009	0.0009	0.0044	0.0027	0.0027	0.0027	0.0009	0.0027	0.0044	0.0027	0.0044	0.0062	0.0009	0.0009	0.0009	0.0027	0.0020
$VAR[E(Ky)]$	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052	0.0052

Table E-3. Estimate of Expected Excess FI Accidents in Last Year (2001) and Its Variance

Subsection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Length (mi)	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.03
Year	FI Accident Frequencies																	
1989	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
–	–	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–
2000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
$\hat{K}_1$	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.004	0.001
$\hat{K}_Y$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.005	0.003	0.003	0.003	0.003	0.001
$E(K_y)$	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.001
Excess	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
$VAR(\hat{K}_Y)$	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	7.10E-06	4.69E-06	4.69E-06	4.69E-06	4.69E-06	1.41E-06
$VAR[E(K_y)]$	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	1.8E-05	5.4E-06

Table E-4. Summed Unweighted and Cost-Weighted Accident Frequency Profiles

Subsection	ES_K for 2001	VAR (ES_K) for 2001	$ERCS_K$ for 2001	VAR ($ERCS_K$) for 2001
1	0.064	0.0356	0.083	1.0673
2	–0.038	0.0302	–0.146	1.0354
3	–0.098	0.0266	–0.206	1.0318
4	0.106	0.0374	0.251	1.0955
5	0.064	0.0356	0.083	1.0673
6	0.004	0.0320	0.022	1.0636
7	0.064	0.0356	0.083	1.0673
8	–0.038	0.0302	–0.146	1.0354
9	0.064	0.0356	0.083	1.0673
10	0.166	0.0410	0.311	1.0992
11	0.124	0.0392	0.143	1.0709
12	0.106	0.0374	0.251	1.0955
13	0.150	0.0391	0.729	1.2200
14	–0.098	0.0266	–0.206	1.0318
15	0.082	0.0375	–0.025	1.0426
16	0.022	0.0338	–0.085	1.0390
17	–0.056	0.0283	–0.038	1.0600
18	0.013	0.0097	0.106	0.3378

Appendix F

Target Accident Types

Target Accident Types

- Multiple-Vehicle Accidents (Manner of Collision)
  - Rear-End
  - Head-On
  - Angle (turning)
  - Sideswipe, Same Direction
  - Sideswipe, Opposite Direction
- Single-Vehicle Accidents
  - Collision with animal
  - Collision with bicycle
  - Collision with parked vehicle
  - Collision with pedestrian
  - Collision with train
  - Overturned
  - Run-off-road
- Light Condition
  - Daylight
  - Dark – street lighted
  - Dark – street not lighted
  - Dark – street lighting unknown
  - Dawn
  - Dusk
- Surface Condition
  - Dry
  - Wet
  - Snow
  - Ice
  - Sand, mud, dirt, oil, gravel
  - Water (standing, moving)
  - Slush
- Crash Severity
  - Property-damage-only (PDO)
  - Nonfatal injury (NFI)
  - Fatal injury (FI)
- Alcohol/Drug Involvement
  - Neither alcohol nor other drugs
  - Yes (alcohol)
  - Yes (drugs other than alcohol)
  - Yes (alcohol and drugs)
- Pedestrian Involvement Indicator
- Bicycle Involvement Indicator