PRE-IMPLEMENTATION EVALUATION OF SAFETY IMPROVEMENT PROGRAMS

by

FRED C. LIN

B.A.Sc. (Civil Engineering), The University of British Columbia, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF CIVIL ENGINEERING

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April, 2000

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Civil Engineering

The University of British Columbia

Vancouver, Canada

Date Feb. 2, 2000

ABSTRACT

This thesis describes the development of an economic-based prediction and evaluation procedure that can be utilized in prioritizing safety improvement projects. The objective of the procedure is to alleviate the problems associated with traditional economic analysis of road safety improvement programs.

Traditionally, the overall effectiveness of safety improvement programs is normally based on the benefits anticipated from a reduction in road collisions. Although the procedures for performing the economic analysis of road safety improvements in general are reasonably straightforward and well documented in the literature, these procedures fail to accurately estimate the safety benefits or disbenefits of these improvements on a consistent basis. The problems can be categorized into two parts: system-wide versus project-level analysis, and dealing with the uncertainties in the effectiveness and applicability of the proposed countermeasure. To resolve these problems, this thesis first describes the development of a new safety analysis software known as ISECR (the Information System for Estimating Crash Reductions) which can be used to determine the expected collision reduction due to a specific countermeasure. ISECR is an intelligent database that uses a case-based reasoning approach and consists of past safety research efforts on collision reduction factors (CRFs) associated with different countermeasures. The system can be used to determine the expected CRFs and the associated range and reliability of the proposed countermeasure when applied to a particular problem at hand.

The safety benefits of implementing a countermeasure at a location can be represented by the expected reduction in collision frequency, which is normally calculated by the product of CRFs and the expected number of collisions. With the ISECR predictions on CRFs and the expected number of collisions determined by the multivariate and Empirical Bayes methods, this thesis then illustrates the use of the moment approach to evaluate the expected collision reduction and its uncertainty. The results can then be used to assist in evaluating the economic feasibility of a countermeasure prior to its implementation. Specifically, the probability of achieving a preset economic goal (i.e., a specific benefit-cost ratio) by implementing a countermeasure at a specific location can be determined.

Finally, the prototype ISECR has been verified and validated using several case studies. The results of the verification and validation have shown that ISECR produced results that are comparable to the results obtained from real cases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would firstly like to express my gratitude to my thesis advisor, Dr. Tarek Sayed, for his constant guidance and supervision throughout the duration of this research. He was very patient, helpful, and provided me with constructive suggestions to this research.

I would also like to thank the Insurance Corporation of British Columbia for providing the financial support for this research and G.D. Hamilton & Associates Consulting LTD. for providing the document summaries used in the ISECR database. Furthermore, I would also like to thank Joan Ng and Vikki Ngan for locating and copying the evaluation studies, and Shaun Bidulka for entering them into the ISECR database.

Most important of all, I would like to specially thank my family for their support and encouragement throughout, not just this research, but my whole life. Their influence and example have inspired me to reach many goals in my life. Thank you, mom and dad, and of course, Frank and Kelly.

Last but not least, special thanks to that special person (you know who you are, Gillian!) for the loving support and most important of all, patience during this endeavor. You are simply the best!

1.0 INTRODUCTION

1.1 Background

In response to limited budgets and growing fiscal constraints, it has become very important to ensure the funding available for road safety improvements is efficiently utilized. Traditionally, funding allocated for road safety improvement programs has been proportionally low compared to other road projects. Nevertheless, due to the increasing public awareness and high social and economic costs of road collisions in recent years, it has become apparent to road agencies responsible for road investment and improvement that the importance of road safety can not be overemphasized. Consequently, it is crucial to improve safety evaluation procedure to ensure an optimal allocation of the available funding.

In an attempt to maximize the overall safety benefit to road users, safety professionals have developed and invoked a standard process to evaluate the cost-effectiveness of road safety projects and programs. Typical safety improvement programs, or commonly referred to as Black Spot Programs, usually include the identification, diagnosis, and remedy of collision-prone locations. In evaluating these programs, the overall effectiveness is normally based on the safety benefits anticipated from a reduction in road collision frequency and/or severity following the implementation of a safety improvement. Procedures for performing the economic analysis of road improvement programs in general are reasonably straightforward and are well documented in the literature. However, the problem with these procedures, when applied to estimating the effectiveness of road safety improvements, is that they do not always accurately estimate the safety benefits of these improvements. It is generally felt by many professionals associated with road safety, that a comprehensive and systematic economic-based approach for the accurate pre-implementation evaluation of road safety programs does not currently exist

1.2 Problem Definition

The problem associated with the traditional economic evaluation procedures arises from the estimation of safety benefits of the proposed improvements. The safety benefits are represented by the expected reduction in the number and/or severity of collisions following the implementation of the improvement. The collision reduction is calculated as the product of the countermeasure effectiveness and the expected number of collisions. The problem can be categorized into two parts: system-wide versus project-level analysis, and dealing with the uncertainties in the effectiveness and applicability of the proposed countermeasure.

A very important step in estimating collision reduction is determining the effectiveness of countermeasures, or what is known in the literature as collision reduction factors (CRFs). A collision reduction factor (CRF) can be considered simply as a value representing the percentage of collisions that a safety improvement is expected to eliminate from a location, or a group of locations receiving the same treatment type.

Several agencies, such as the Federal Highway Administration (FHWA) and the Institute of Transportation Engineers (ITE), have developed CRFs for different safety improvements. Despite the common usage of these CRFs in practice, there are two major deficiencies in using them to generate project-level safety estimates. Firstly, these CRFs are developed as system-wide factors where the variability of collision patterns and geometric configuration among different locations are not considered. Secondly, the use of these CRFs does not account for the treatment of the uncertainty issues related to the effectiveness of the proposed safety improvement. Means to resolve these two problems will be presented in this thesis.

1.3 Thesis Objectives

The objective of this thesis is to present a systematic procedure that aims to assist in prioritizing safety improvement projects and ultimately ensure the optimal allocation of funding available for road safety improvements. The followings are the key focal points:

1. Investigate methods to evaluate the effectiveness (CRFs) of various safety improvements on a project-level basis where different collision patterns and site characteristics of a location are considered.
2. Explore techniques to deal with the uncertainty issues involved in determining the effectiveness (CRFs) of various safety improvements. Specifically, the expected values of CRFs and the associated standard deviations will be calculated.
3. Present an economic-based prediction and evaluation procedure to increase the confidence of predictions associated with the current practice of evaluating the expected safety benefits from improvement programs. Specifically, the probability of a proposed safety improvement achieving a specific economic goal (e.g., a pre-determined benefit-cost ratio) prior to its implementation will be calculated.

1.4 Methodology

To estimate CRFs on a project-level basis, a new safety analysis software called the Information System for Estimating Crash Reductions (ISECR) is developed in this thesis. ISECR is an intelligent database that uses case-based reasoning and consists of past safety research efforts on CRFs associated with the different safety improvement measures. ISECR has the following functions:

1. Data entry: ISECR permits the entry of evaluation studies that report CRFs associated with different safety improvements.
2. Data retrieval and analysis: ISECR employs a case-based reasoning (CBR) approach for data retrieval and analysis. The system is designed to accept queries, analyze and display information from the database that matches the queries.
3. CRFs estimation and reporting: ISECR computes the effectiveness and the associated range and reliability of the proposed countermeasure related to the total number of collisions, and different collision types and severity.
4. Benefit-cost ratio estimation: ISECR provides results that can be used in evaluating the benefit-cost ratio pertaining to the proposed safety improvement.

1.5 Thesis Structure

This thesis is divided into nine chapters. Following this introductory chapter, Chapter Two presents the results of a comprehensive literature review on the topics of case-based reasoning and collision analysis. Chapter Three explains how CBR is utilized in the design of the intelligent database, ISECR. As well, Chapter Three details the methodology which ISECR uses to estimate CRFs and their uncertainties. Chapter Four describes the various functions provided by ISECR. The validation of SECR is presented in Chapter Five. In Chapters Six and Seven, the uncertainty issues involving the expected reduction in collision frequency and economic analysis are examined respectively. An example illustrating the procedures outlined in this thesis is included in Chapter Eight. Finally, the conclusion and some suggestions for further research are included in Chapter Nine.

2.0 LITERATURE REVIEW

2.1 Introduction

This chapter presents the results of a comprehensive literature review on the following two topics:

1. Case-base reasoning (CBR)
2. Collision analysis, with a focus on the generalized linear modeling (GLIM) approach and the Empirical Bayes (EB) technique used to determine the expected number of collisions at a location.

The first part of this chapter provides a literature review on CBR, which is utilized in the development of ISECR to predict collision reduction factors (CRFs) and the associated uncertainties. As noted in Chapter One, most of collision analyses utilize the products of CRFs and the expected number of collisions to estimate the potential safety benefits, i.e., the expected collision reduction, of implementing a safety improvement at a location. Thus, the second part of this chapter focuses on describing how the GLIM and EB approaches can be used to estimate the expected number of collisions and the associated uncertainties at a location.

2.2 Case-Based Reasoning

Knowledge-based systems (KBS) utilize rule-based and model-based reasoning techniques to build design automation and design decision support systems. Although there was notable success in some areas, difficulties have been encountered with KBS when formalizing generalized design experiences such as rules, logic, and domain models. These in turn, have led to the development of case-based reasoning (CBR) systems.

CBR systems have proven to be effective in providing desired results in various fields, such as in architectural design, structural design, software specifications design, and etc. CBR is a relatively recent concept, with the original research conducted in the early 1980s as first described conceptually by Schank (1982). It is a methodology used to solve new problems by reusing and adapting (or combining) solutions that worked for similar problems in the past.

In CBR, past problem solving experiences (cases) are stored in a database (case base) and upon request, cases similar and/or relevant to the current problem (design case) can be retrieved from the case base. The cases retrieved (retrieved cases) from the case base can then be adapted or combined to better fit the design case. Differentiating from the more traditional rule-based and model-based reasoning techniques, i.e., the former captures knowledge in the form of if-then rules while the latter formulates knowledge in the form of principles and/or models to cover various aspects of a problem domain, CBR is an experience-based method that utilizes prior problem solving experiences as its main knowledge source.

CBR is a growing field as it resembles, to some degree, the psychological process at which a person follows when attempting to utilize his/her knowledge and experience in solving a new problem. Using the CBR approach, modeling of domains that are not completely understood and open-ended is possible. Consequently, CBR can be considered as an alternative to paradigms such as rule-based and model-based reasoning.

2.2.1 CBR Applications in Transportation Engineering

In the transportation field, there are few applications that employ a CBR approach. For example, Khattak and Kanafani (1996) first developed a planning tool for ITS (Intelligent Transportation Systems) using a CBR approach. As an enhanced system for their first proposed CBR planning tool, Khattak and Renski (1999) then focused their effort on planning HOV (High Occupancy Vehicle) lanes by integrating the CBR approach in a GIS (Geographic Information System) environment.

In the field of road safety analysis, only one system was found in the literature that employed a CBR approach. Capus and Tourigny (1998) developed a CBR system (called ROSAC, ROad Safety Analysis with Cases) capable of retrieving similar cases from the case base. Furthermore, ROSAC can reuse, adapt, and save the new problem and adapted solution as a new case in the case base. In ROSAC, each case consists of a problem, as represented by site characteristics and collision statistics, and its solution, as described by collision patterns, collision causes, and the implemented safety improvements. Case retrieval is realized when the site characteristics and collision statistics of the current situation match, to some degree, with the ones stored in the case base. The retrieved cases can then be reused and adapted in ROSAC as the solution to the design case.

The following sections provide a brief introduction to some of the issues and procedures involved in a CBR system design, with a focus on the techniques utilized in the development of ISECR. However, for a detailed and comprehensive description on CBR, books and/or reports written by authors such as Kolodner (1993), Watson (1997), Leake and Plaza (1997), etc., should be consulted.

2.2.2 General CBR Issues and Procedures

Regardless of the domain of application, the followings are the two general issues concerning the application and design of any CBR system:

1. Representation issues: Representation issues concern with the contents (information), representation, indexing, and memory organization of cases.
2. Control issues: Control issues, on the other hand, deal with the general processes of a CBR system, which are comprised of the retrieval, adaptation, and combination of cases.

Depending on the context of each individual project, the representation and control issues involved in a CBR system design may vary from one project to another. Nevertheless, the main purpose of any CBR system is to facilitate the solving of a similar problem in a somewhat similar context. Hence, general design issues and procedures involved in all CBR systems can be observed and are further discussed below.

2.2.3 Representation Issues of CBR

In any CBR system, the database (case base) contains a representation of a set of previously solved problems (cases). It is extremely important to select an appropriate model for case representation when designing a CBR system. An appropriate model provides the very basis of how cases are represented in the CBR case base. Furthermore, the subsequent CBR procedures, including the retrieval of relevant cases and their adaptation or combination, rely heavily on the representation model selected. For that reason, a systematic approach is required to identify and express the uniform representation of specific features that make up a case.

Case representation, as mentioned previously, is the first of two issues that needs to be addressed prior to carrying out any CBR system design. In general, issues related to developing an adequate and useful case base can be addressed by answering the following four fundamental questions:

1. What are the contents, or features, that are essential to represent a case?
2. How to express or represent these features (as in numbers, symbols, Boolean variables, texts, models, etc.)?
3. How to structure these features effectively and efficiently to minimize the storage and computation requirements?
4. What case indexing schemes should be used for the retrieval of cases?

Essentially, case representation is concerned with how the contents of each case are represented and organized in the case base. There are a number of ways to organize the information in a case, however, only two of these alternatives are relevant to this research and discussed further below, i.e., using a set of attribute-value pairs or part-subpart relationships. These two alternatives are illustrated in Figure 2.1.

When representing a case as attribute-value pairs, all features and associated values are represented with these sets of values. Information gathering is made simple in this manner as all aspects of a case are kept in one case. This type of case representation is typically adequate for small and simple cases. When representing a case as a hierarchy of part-subpart relationships, information within a case is broken down into a number of subcases. Therefore, a higher degree of information is required to adequately represent the subcases, and to designate the relationship knowledge within a case. Despite the higher degree of complexity with this type of case representation when compared to the attribute-value pairs, using a part-subpart hierarchy is beneficial as it facilitates the representation of large and complex cases.

The representation and organization of a case should be consistent for all cases in the case base. In this regard, all cases are described by the same set of features, represented either by attribute-value pairs and/or part-subpart relationships. An adequate representation scheme is the first step to ensure a successful CBR system design.

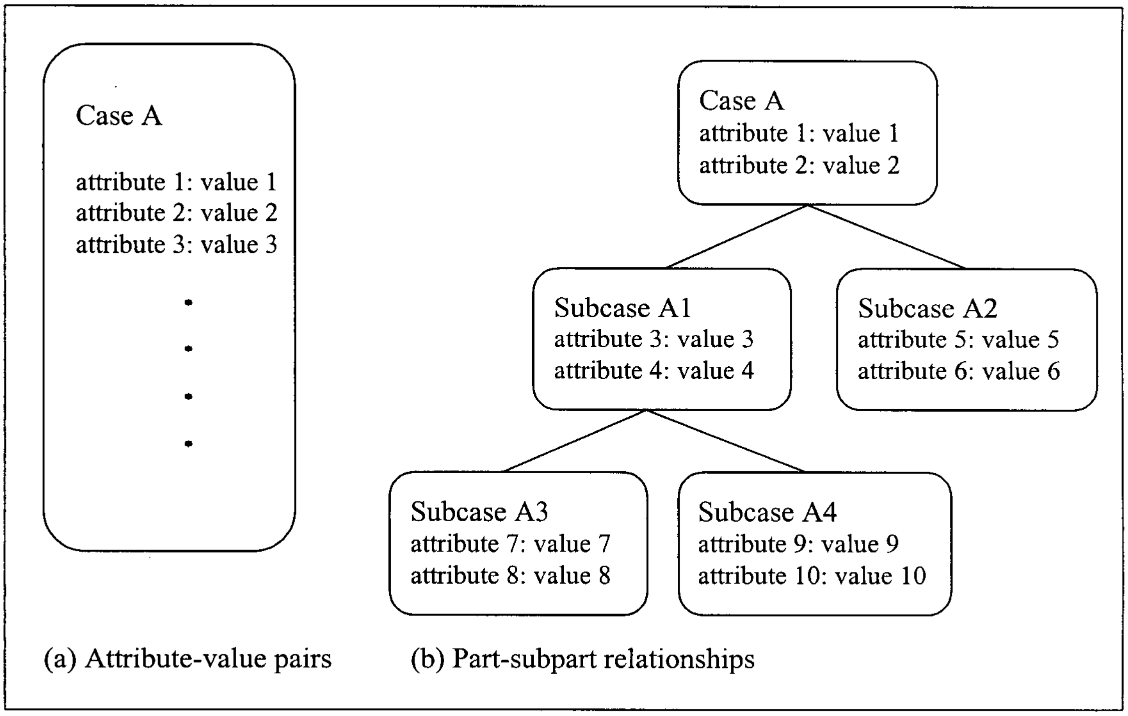


Figure 2.1. Representation of a Case in Case Base

2.2.4 Control Issues of CBR

Once the above case representation issues are addressed, the following three general CBR processes, as often referred to as the control issues, can then be carried out sequentially:

1. Case retrieval,
2. Case adaptation, and
3. Case combination.

Case retrieval refers to the capability of a CBR system to accept queries and filters information from the case base which matches the queries. It is uncommon that the features in a design case match completely with the ones in the retrieved cases. The simplest approach would be to retrieve and reuse the unchanged solution of the most similar case as the solution to the design case. While this simplistic approach reduces the computation and programming requirements needed to design a CBR system, it fails to account for the differences between the design and retrieved cases. Since no two problems are ever the same, it is necessary for any CBR system to have at least one of the adaptation or combination capability (Pu, 1998). Solution to the design case can then be constructed either by adapting or combining old solutions. In this manner, use of a CBR system is more advantageous as the adaptation or combination of previously solved solutions is easier than generating a new solution from scratch.

2.2.4.1 Case Retrieval

Case retrieval is a basic operation in any CBR design, but it plays a significant role in the establishment of a CBR system. Retrieval of cases can be done informally or formally. Informal case retrieval refers to the selecting of relevant cases from the case base by the user based on his/her experience and/or judgement. On the other hand, formal case retrieval refers to a CBR design system capable of accepting a new set of definition and/or features from the user and conducting searches on its case base for cases that have the same or similar problem specifications. ISECR utilizes the latter of the two retrieving techniques.

Retrieval of cases can be based on a perfect match, where the specifications of the design case are found to match exactly. However, it is more likely to find the features in the design case that do not match completely with the ones in the retrieved cases. Thus, most of the retrieved cases from a CBR system are partial matches.

The degree of partial match depends on the relative similarity and importance of the features that make up a case. The degree of similarity of each feature can be calculated by comparing the specific feature in the design case to the corresponding feature of retrieved cases. Case distance can be computed and used as a measure to determine the degree of similarity between the design and retrieved cases. One of the most frequently used case distances is the Euclidean distance calculated as (Yeh, 1997):

$$D_k = \left[\frac{\sum_{i=1}^{n} w_i \cdot ABS\left(\frac{f_i^I - f_{i,k}^R}{f_{i,\max}^R - f_{i,\min}^R} \right)^2}{\sum_{i=1}^{n} w_i} \right]^{\frac{1}{2}} \qquad \textit{(Equation 2.1)}$$

where D_k = Case distance of the design case to the k^{th} case

n = Number of input features

w_i = Importance factor for feature i

ABS = Absolute value function

$f_{i,\max}^R$ = Maximum value for feature i

$f_{i,\min}^R$ = Minimum value for feature i

f_i^I = Value of feature i in the design case

$f_{i,k}^R$ = Value for feature i in the k^{th} case

In Equation 2.1, the importance factor, w_i can be represented by a numerical value ranging from 0.0 to 1.0. A higher importance value (closer to 1.0) indicates that the feature is more important compared to the features having lower importance values. The value of case distance, D_k, as determined by Equation 2.1, can also be any real number varying between 0.0 to 1.0. The more similar the design case to the retrieved case, the smaller the D_k value or shorter the case distance. Case distance can be used to rank cases, i.e., lower scoring cases are used before higher scoring ones. Inherently, cases are ranked based on a weighted sum of features in the design case that match the ones in the retrieved cases.

2.2.4.2 Case Adaptation

The main purpose of case adaptation is to modify the solution of the case retrieved from the case base to account for the differences between the design and retrieved cases. There are a number of adaptation methods available, i.e., reinstantiation, parameter adjustment, local search, case-based substitution, commonsense transformation, model-guided repair, special-purpose adaptation and repair, derivational reply, etc. In general, adaptation methods in CBR are classified into the following two categories:

1. Structural adaptation: The retrieved case solution is substituted directly with adaptation rules and/or formulas to generate a new solution for the design case.
2. Derivational reply: The same methods at which the original case solution (from the retrieved case) is derived are reused to derive a new solution for the design case. Specifically, it is assumed that the same rules and/or formulas used to generate the retrieved case solution can be reused to produce the new solution to the design case. With this method, the sequence at

which the retrieved case is solved must be stored in the case base as an additional attribute to the case.

Although adaptation of cases is useful in many applications, many successful CBR systems do not perform adaptation at all or leave this option to the user. Watson (1997) suggested that adaptation should be avoided unless it can be carried out easily using simple adaptation methods such as reinstantiation or parameter adjustment. Complex adaptation of cases is knowledge intensive and can only be applied to domains that are well understood. Since CBR is generally employed to solve problems that are not well understood, a complex adaptation method that is knowledge intensive may not be feasible.

2.2.4.3 Case Combination

Case combination aims to derive the solution for the design case by combining several of the original solutions stored in the retrieved cases. Frequently, the solution stored in the most similar case may not be the best solution for the design case. Thus, it becomes necessary to retrieve several similar cases and combine their solutions in order to generate an improved solution to the current problem (Yeh, 1997). Case combination can be carried out in several ways, such as the weighted average approach, constructive approach, or frame approach (Pu, 1993, and Yeh, 1997). However, for the purpose of this research, the weighted average approach is adequate to the application and it is shown below:

$$T_{Weighted} = \frac{\sum_{k=1}^{N}(1 - D_k) \cdot T_k}{\sum_{k=1}^{N}(1 - D_k)} \qquad \textit{(Equation 2.2)}$$

where $T_{Weighted}$ = Weighted solution for the design case

N = Number of retrieved cases used to generate the weighted solution

D_k = Case distance of the design case to the k^{th} case, as determined by Equation 2.1

T_k = Solution for the k^{th} case

The term $(1-D_k)$ in Equation 2.2 suggests that solutions obtained from cases with lower case distances, i.e., more similar to the design case, are weighted more significantly compared to the ones with higher case distances. The accuracy of the above weighted solution can be represented by the standard deviation that has the following expression:

$$Stdev(T_{Weighted}) = \sqrt{\frac{\sum_{k=1}^{N}(1-D_k)\cdot(T_k - T)^2}{\sum_{k=1}^{N}(1-D_k)\cdot(N-1)}} \qquad \textit{(Equation 2.3)}$$

where $Stdev(T_{Weighted})$ = Standard deviation of the weighted solution for the design case

2.2.5 CBR Methodologies

There are several CBR methodologies used in practice, each employing different combination of the three CBR processes described in Sections 2.2.4.1 to 2.2.4.3. In this research, two of the CBR methodologies are utilized in the design of ISECR and thus, discussed further below.

2.2.5.1 Nearest Neighbour Approach

The first CBR methodology is called the nearest neighbour approach and is illustrated in Figure 2.2. This is the simplest of all CBR methodologies where the only CBR process used is case retrieval. With this approach, only the most similar case from the case base is retrieved and the result stored in this retrieved case is then utilized as the solution to the design problem.

2.2.5.2 Collaborative Approach

Moving to a more sophisticated level, the second CBR methodology is called the collaborative approach as shown in Figure 2.3. Collaborative CBR approach retrieves several cases from the case base and performs case combination to the retrieved cases. With this technique, k-nearest neighbour retrieval is used, where k refers to a predetermined number of cases to be retrieved. The retrieved case solutions are then combined (using Equation 2.2) to yield the solution for the design case. The accuracy of the solution generated for the design case can be estimated by calculating its standard deviation by using Equation 2.3. Essentially, collaborative CBR approach utilizes both case retrieval and case combination to derive a better-fit solution to the design case.

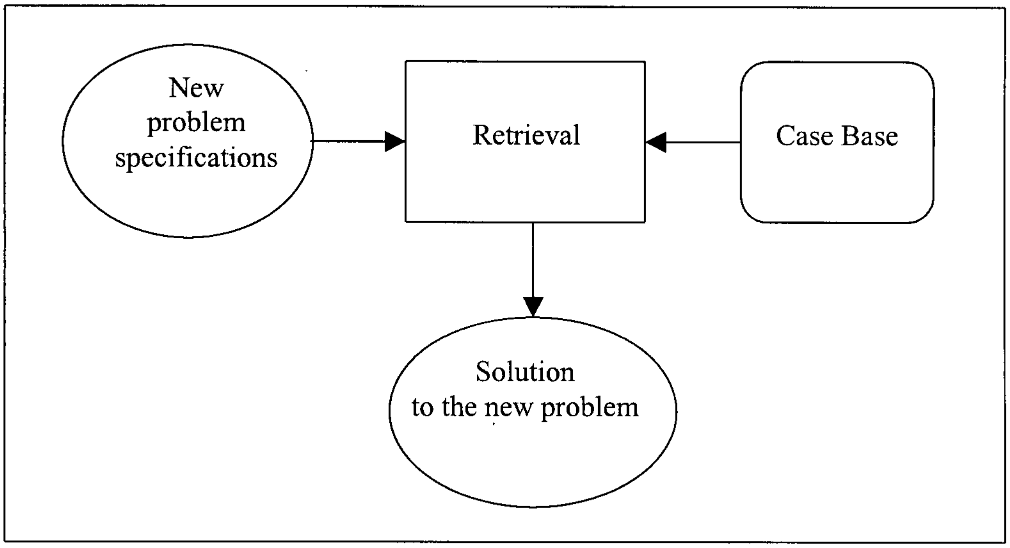


Figure 2.2. Nearest Neighbour Approach

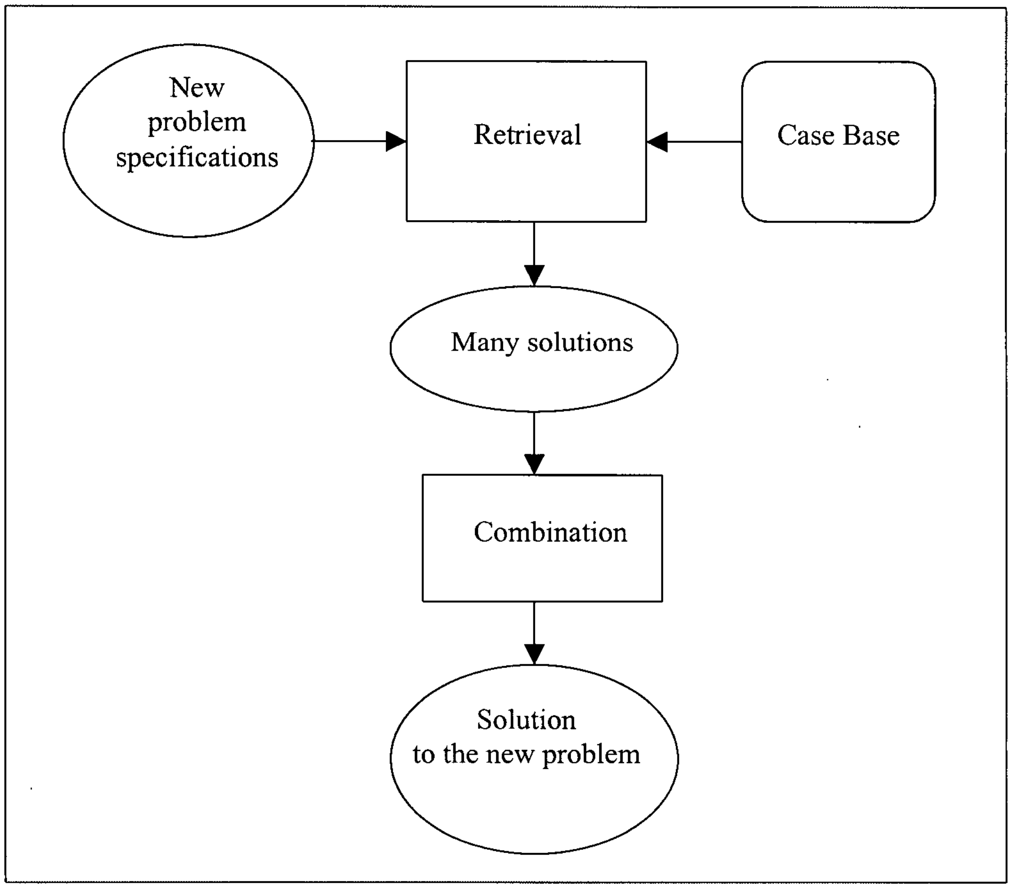


Figure 2.3. Collaborative Approach

2.3 Collision Analysis

The need for using the generalized linear regression modeling (GLIM) and the Empirical Bayes (EB) approaches arises as they address and overcome the problems associated with the conventional methods of predicting site-specific collision estimates. The following sections review some of the published research to construct a fundamental understanding on the above two techniques.

2.3.1 The Generalized Linear Modeling (GLIM) Approach

Most of the earlier work in collision analysis utilized the conventional linear regression approach to develop prediction models relating collisions to traffic volumes with the assumption of a Gaussian (normal) distributed error structure. However, several researchers (Jovanis and Chang, 1986, Hauer et al., 1988, Saccomanno and Buyco 1988, Miaou and Lum 1993) have all shown that conventional linear regression prediction models lack the distributional property to adequately describe the random, discrete, non-negative, and typically sporadic characteristics of traffic collisions. To overcome these problems, Jovanis and Chang (1986) concluded, with supporting results from their modeling of collisions at highway sections in Indiana, that a Poisson distribution should be used to describe the model error structure. Furthermore, Miaou and Lum (1993) also supported the use of a Poisson distributed error structure with their results. All in all, these researchers confirmed the shortcomings associated with the conventional linear regression models and their capabilities in predicting collision estimates.

The GLIM modeling approach, on the other hand, can overcome the shortcomings associated with the conventional methods. Specifically, utilizing the GLIM approach, the flexibility of assuming different error structures is provided and the conversion of non-linear models into linear ones is also feasible. As an example, the followings describe how Hauer et al. (1988) and Kulmala (1995) utilized the GLIM approach to develop collision prediction models for intersections.

Assuming that Y is a random variable that describes the number of collisions in a specific time period, y is the observation of Y during a period of time, and the random variable Λ is regarded as the mean of Y. Thus, for $\Lambda = \lambda$, Y is Poisson distributed with parameter λ. If each site has its own regional characteristics with a unique mean collision frequency Λ, Hauer et al. (1988) have shown that for an imaginary group of sites having similar characteristics, Λ follows a gamma distribution. The gamma distribution, having parameters κ and κ/μ, has a mean and variance that can be described with the following two equations:

$$E(\Lambda) = \mu \qquad \text{(Equation 2.4)}$$

$$Var(\Lambda) = \frac{\mu^2}{\kappa} \qquad \text{(Equation 2.5)}$$

Based on Equations 2.4 and 2.5, the point probability function of Y can be given by the negative binomial distribution with an expected value and variance described by Equations 2.6 and 2.7 respectively as shown below (Hauer et al. 1988, and Kulmala, 1995):

$$E(Y) = \mu \qquad \textit{(Equation 2.6)}$$

$$Var(Y) = \mu + \frac{\mu^2}{\kappa} \qquad \textit{(Equation 2.7)}$$

As shown above, unless $\kappa \to \infty$, the variance of the observed number of collisions is greater than its expected value. However, when $\kappa \to \infty$, the variance equals the expected value, which is identical to the Poisson distribution (Kulmala, 1995) and suitable to describe the nature of collisions.

2.3.2 The Empirical Bayes (EB) Approach

The EB approach is used to refine the GLIM estimate of the expected number of collisions at a location to yield a more accurate, location-specific safety estimate. Two types of clues of the location are used in the EB approach: its traffic and road characteristics, and its historical collision data (Hauer, 1992, Brude and Larsson, 1988). Utilizing the EB approach, Hauer (1992) calculated the expected number of collisions for any intersection using the following equations:

$$EB_{Safety.Estimate} = \alpha \cdot E(\Lambda) + (1-\alpha) \cdot count \qquad \textit{(Equation 2.8)}$$

where $\alpha = \dfrac{1}{1 + \dfrac{Var(E(\Lambda))}{E(\Lambda)}}$ *(Equation 2.9)*

$count$ = observed number of collisions

$E(\Lambda)$ = predicted number of collisions estimated from the GLIM model

$Var(E(\Lambda))$ = variance of the GLIM estimate

Since $Var(E(\Lambda)) = \frac{E(\Lambda)^2}{\kappa}$, Equation 2.8 can be rearranged as:

$$EB_{Safety.Estimate} = \left(\frac{E(\Lambda)}{\kappa + E(\Lambda)}\right) \cdot (\kappa + count) \qquad \textit{(Equation 2.10)}$$

Finally, the variance of the EB estimate can be determined using Equation 2.11 as:

$$Var(EB_{Safety.Estimate}) = \left(\frac{E(\Lambda)}{\kappa + E(\Lambda)}\right)^2 \cdot (\kappa + count) \qquad \textit{(Equation 2.11)}$$

Hence, the safety estimate, i.e., the expected number of collisions and the associated uncertainties, can be evaluated by using Equations 2.10 and 2.11 respectively.

3.0 CBR APPLICATION IN ISECR

3.1 Introduction

The computerized approach of the Information System for Estimating Crash Reductions (ISECR) is designed to minimize the amount of manual work required in evaluating the effectiveness of different safety improvements on a project-level basis. ISECR consists of historical information extracted from past evaluation studies that reported the performance of different safety improvements. Utilizing a CBR approach, ISECR is capable of retrieving and analyzing appropriate past records in order to assess the range and reliability of the predicted countermeasure effectiveness (CRFs) for a given safety improvement under certain condition. This chapter provides an overview of the implementation of CBR in ISECR by focusing on the followings:

1. Case representation: Determine the contents/features used to represent a case, and their organization and structure in the ISECR case base.
2. Case retrieval: Establish the query parameters used to retrieve relevant cases.
3. Case distance: Determine the criteria used to calculate the case distance for each of the retrieved cases.
4. Solution construction: Establish the adaptation and/or combination strategies employed to modify and/or combine the retrieved solutions to create an improved solution for the design case.

3.2 Case Representation in ISECR

Frequently, more than one CRF can be found in an evaluation study, i.e., some studies reported the effectiveness for more than one countermeasure while others evaluated the same countermeasure under different situations, etc. Hence, it is possible to represent an evaluation study with more than one case in the ISECR case base. The information contained within each case is organized using part-subpart relationships, as described in Section 2.2.3. The contents within each case are broken down into the following six subcases:

1. General information,
2. Case quality,
3. Countermeasure type,
4. Location type,
5. Location characteristics, and
6. Case solutions.

Within each subcase, attribute-value pairs are used to represent the features and associated values. The representation of each of the above six subcases is organized as follows and shown in Figure 3.1:

Subcase 1- General Information:

- Case Id: reference point for a given case in the case base
- General information on the evaluation study of which the case belongs to:

- Study Id: an identifier used to reference the evaluation study
- Author(s) of the study
- Study title
- Source of the study: journal, volume, pages, publication date, and country

Subcase 2- Case Quality:

- This subcase stores information regarding the treatment of the following three confounding factors:
  - Changes in traffic volume
  - Inclusion of unrelated effects
  - Regression to the mean (RTM) artifact

Subcase 3- Countermeasure Type:

- A total of 116 countermeasure types are considered in this subcase (see Appendix A for details). They are organized and grouped into the following sixteen countermeasure categories: area-wide schemes, bridge improvements, cyclist/pedestrian facilities, delineation, geometric improvements, intersection improvements, lane/shoulder treatment, lighting improvements, object removal/relocation, parking improvements, pavement treatment, railway improvements, regulation change, safety barriers, traffic controls/signs, and traffic signals.

Subcase 4- Location Type:

- This subcase gathers the information on the type of location where the countermeasure effectiveness is investigated. A total of eight location types are considered in this research, i.e., general intersections, signalized intersections, unsignalized intersections, road sections, freeways, bridges, rails, and construction zones.

Subcase 5- Location Characteristics:

- The location characteristics associated with a given case are extracted from the evaluation study and stored in this subcase. Different location characteristics are considered for each of the eight location types (see Appendix B for details). For example, the following characteristics are considered for signalized intersections (see Table 3.1):

1. Area type: urban, suburban, rural, or other
2. Intersection type: four-legged, t-intersection, or y-intersection
3. Implementation level: isolated location or wide area
4. Total traffic volume: Total entering traffic volume (AADT) recorded for the intersection. The intervals are as follows: 0-4999, 5000-9999, 10000-14999, 15000-19999, 20000-29999, 30000-39999, 40000-49999, 50000-59999, 60000-69999, 70000-79999, and 80000 and more.
5. Average lane width: Average width of all traffic lanes. The distinction is made between less than 12 feet and greater or equal to 12 feet.
6. Provision of left-turn channelization: yes or no

7. Provision of right-turn channelization: yes or no
8. Left-turn movement: not allowed, permissive, or protected
9. Right-turn movement: not allowed, permissive, or protected
10. Street parking: allowed or not allowed
11. Type of traffic control: fixed-timed, semi-actuated, or fully-actuated
12. Average number of lanes per approach: less than or equal to two lanes or more than two lanes

- The above location characteristics are represented in the case base as different sets of attribute-value pairs. For example, Table 3.1 presents a set of attribute-value pairs to represent location characteristics for signalized intersections (see Appendix B for the representation of the location characteristics used for other location types). Representing these features in such a manner does not only minimize the storage requirement, but also facilitates the calculation of case distance, as will be shown later.

Table 3.1. Representation of Location Characteristics for Signalized Intersections

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Intersection type	1=four-legged 2=t-intersection 3=y-intersection
Implementation level	1=isolated location 2=wide area
Total entering traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Provision of left-turn channelization	1=yes 2=no
Provision of right-turn channelization	1=yes 2=no
Left-turn movement	1=not allowed 2=permissive 3=protected
Right-turn movement	1=not allowed 2=permissive 3=protected
Street parking	1=allowed 2=not allowed
Type of traffic control	1=fixed-timed 2=semi-actuated 3=fully-actuated
Average number of lanes per approach	1=less than or equal to 2 lanes 2=more than 2 lanes

Subcase 6- Case Solutions:

- The reported collision reduction factors (CRFs) are entered into this subcase as the case solutions. CRFs corresponding to the following collision severity and types are entered into the case base (if available):
  - Total collisions
  - Collision severity: fatal, injury, casualty, and property-damage-only.
  - Collision types: angle, bike/pedestrian, fixed-object, head-on, left-turn, off-road, overtaking, parked vehicle, rear-end, right-turn, sideswipe, and other.

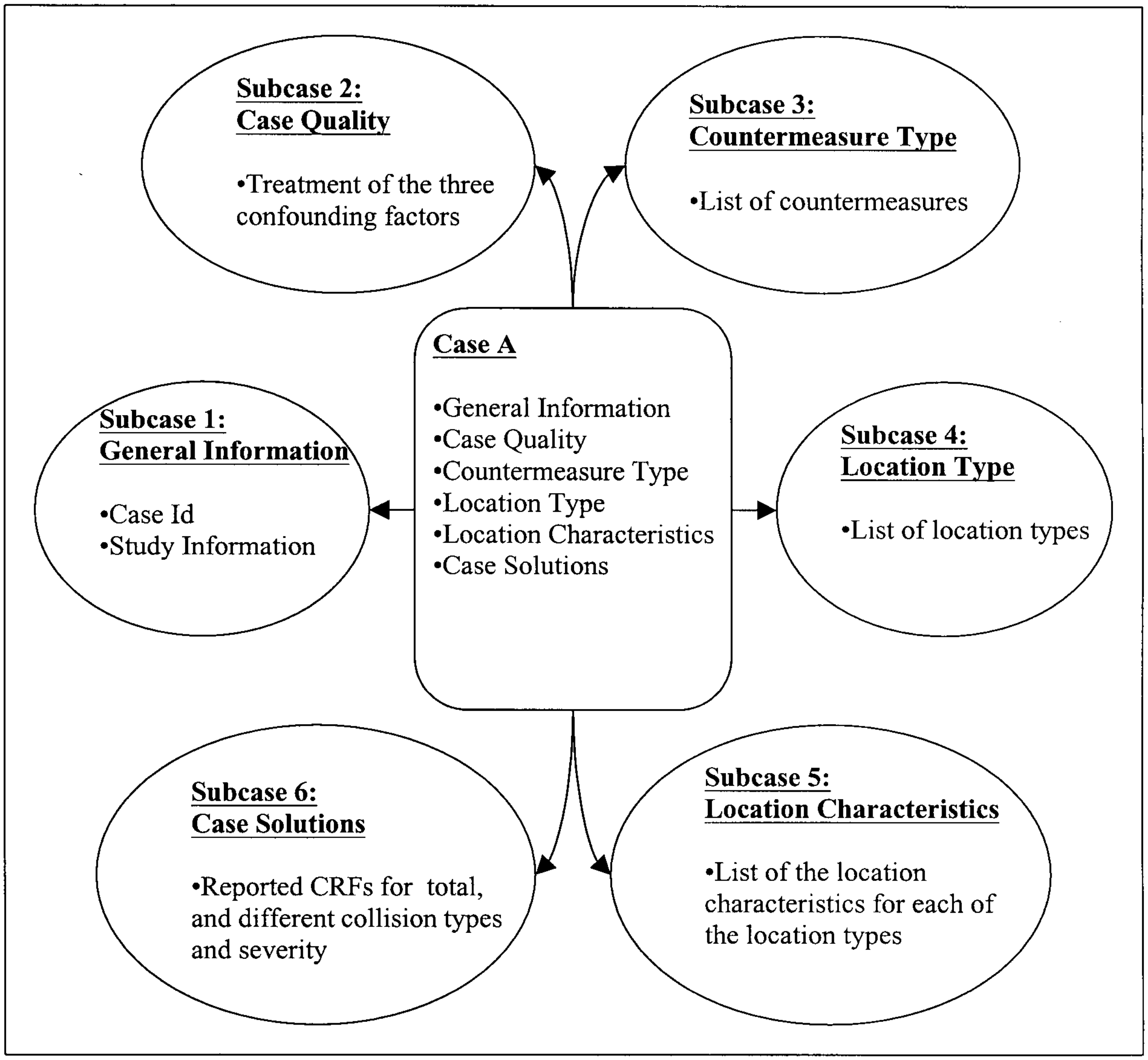


Figure 3.1. Representation of a Case in ISECR

With the above information entered into the case base, the ISECR user can then query for cases that report the effectiveness of a specific countermeasure at a specific location. Case retrieval strategies used in ISECR are further explained in the next section.

3.3 Case Retrieval in ISECR

Case retrieval in any CBR system requires the user to input a set of definitions and/or features, which can be used to uniquely describe the current problem (design case). These specifications become the query parameters for filtering cases from the case base. When using ISECR, the user is first asked to specify both the type and location of the proposed countermeasure by selecting from the same lists used in Subcases 3 and 4 respectively (see Section 3.2). These two features are then used as the query parameters and are matched against the associated attribute-value pairs contained in each case stored in the case base. A case is only retrieved from the case base if a perfect match of the two query parameters is found.

It is likely that more than one case with varying CRFs will match the above query criteria and will be retrieved from the case base. In this instance, it is essential to determine which of the retrieved cases and the associated results are more valid and/or relevant to the design case. As mentioned previously, a credibility factor and an application factor, or referred to in this thesis as the quality and relevance scores respectively, can be evaluated for each case to assess its validity and relevance.

The following sections discuss the quality and relevance scores in greater depth, the methods in which they are derived, followed by an explanation on how solution construction is implemented in ISECR.

3.4 Quality Score

Quality score is a quantitative measure used to determine the validity of a case and its results. In this thesis, the quality of a case is influenced by the treatment of the three common confounding factors, as considered in Subcase 2 (see Section 3.2) of each case. Confounding factors are factors that may affect the accuracy of the evaluation on the effectiveness of a safety improvement if they are not accounted for in the retrieved case. The lack of treatment of these factors, in turn, can threaten the validity of cases and their results. Ultimately, by treating these confounding factors, one can decide if the observed changes in road safety are caused by the implemented countermeasure, the existence of the confounding factors, or a combination of both.

Some researchers have attempted to account for the presence of confounding factors in evaluation studies. One approach is to assign arbitrary weights to evaluation studies based on their treatment of these factors. For example, Elvik has assessed the quality of evaluation studies based on several confounding variables (Elvik, 1995, 1996, and 1998). These variables include research design, decade of study, change in traffic volume, regression to the mean, collision migration, etc. In this thesis, the following three confounding factors are considered:

1. Changes in traffic volume,
2. Inclusion of unrelated effects, and
3. Regression to the mean (RTM) artifact.

3.4.1 Changes in Traffic Volume

Changes in traffic volume are usually controlled by calculating collision rates (Elvik, 1996). Collision rates can be expressed as collisions per million-entering-vehicles (col/mev) for intersections and collisions per million-vehicle-kilometers (col/mvk) for road sections. Changes in collision rate are often used as a measure of the effectiveness of a countermeasure. By using collision rates, the number of collision is assumed to relate linearly with traffic volumes. However, this assumption is not always valid. In fact, a non-linear relationship between collision frequency and traffic volumes has been shown to be a more suitable assumption (Hauer et al., 1988). This indicates that the use of collision rates does not necessarily, under all circumstances, minimize the effect of traffic volumes on collision frequency. Nevertheless, in this thesis, similar to Elvik's approach (1996), cases that employ changes in collision rate as the measure of countermeasure effect are classified as having accounted for the confounding factor on the changes in traffic volume.

3.4.2 Inclusion of Unrelated Effects

Frequently, factors other than the treatment may affect the observed difference in collision frequency. The inclusion of unrelated effects in a case may lead one to believe that the implemented countermeasure is more effective than it really is. The presence of unrelated effects in a case can be controlled by using a comparison group when determining the CRF. Inherently, a comparison group is a group of sites that are somewhat similar to the treatment site. This group often consists of the total number of collisions in the area where the treatment area is located. The change in collision frequency of the comparison group can be compared to the one observed for the treatment site. This permits the calculation of the actual treatment effects. Therefore, in this research, if a case includes a comparison group when estimating the countermeasure effectiveness, it is classified as having treated for the confounding factor of including the unrelated effects.

3.4.3 Regression to the Mean (RTM) Artifact

Sites are usually selected for treatment due to their high collision occurrence. However, this high occurrence may be entirely caused by a random up-fluctuation of collision around the location's true mean (collision occurrence) value. If a location is selected for treatment solely because it undergoes an up-fluctuation in collision frequency, it will show a reduction in collision occurrence in the after period regardless of the implementation of the countermeasure. Hence, if the RTM bias is not accounted for in a case, an overestimation of the effectiveness of a countermeasure can take place.

RTM artifact can be controlled with the use of a reference group and/or an appropriate analysis technique. A reference group should be selected to represent the treatment site, i.e., from the same potential treatment population. As for the analysis methods used to address the RTM bias, the Empirical Bayes technique is often used in practice. Thus, if a case uses one of these two methods to remove the RTM effects, it is considered to have accounted for this confounding factor.

3.4.4 Calculation of Quality Scores

In ISECR, quality scores are calculated automatically using the information contained in Subcase 2, i.e., case quality, and hence, not requiring any input from the user. Quality score, employing the Euclidean distance, can be computed using Equation 3.1 as shown below:

$$D_{k(Quality)} = \left[\frac{\sum_{i=1}^{n} w_i \cdot ABS\left(\frac{f_i^I - f_{i,k}^R}{f_{i,\max}^R - f_{i,\min}^R} \right)^2}{\sum_{i=1}^{n} w_i} \right]^{\frac{1}{2}} \qquad \textit{(Equation 3-1)}$$

where $D_{k(Quality)}$ = Quality score of the k^{th} retrieved case

n = Number of input features

w_i = Importance factor for feature i

ABS = Absolute value function

$f_{i,\max}^R$ = Maximum value for feature i

$f_{i,\min}^R$ = Minimum value for feature i

f_i^I = Value of feature i in the design case

$f_{i,k}^R$ = Value for feature i in the k^{th} retrieved case

In Equation 3.1, values of 3.0 and 0.0 are assigned to $f_{i,\max}^R$ and $f_{i,\min}^R$ as the maximum and minimum numbers of confounding factors that can be accounted for in a case respectively. A feature value, $f^R_{i,k}$, is assigned to each case based on its treatment of the three confounding factors. For example, if two of the three confounding factors are accounted for in a case, a feature value of 2.0 is assigned. Similarly, if a case has accounted for none, one, or all of the confounding factors, a feature value of 0.0, 1.0, or 3.0 is given respectively. In Equation 3.1, a value of 1.0 is assigned to w_i as the only feature considered at hand is the treatment of the three confounding factors. Lastly, a value of 3.0 is assigned to f^I_i as the optimum number of confounding factors that can be accounted for in any case.

Below is a sample calculation of the quality score (using Equation 3.1) for a case having accounted for two of the three confounding factors:

$$D_{k(Quality)} = \left[\frac{\sum_{i=1}^{n} w_i \cdot ABS\left(\frac{f^I_i - f^R_{i,k}}{f^R_{i,\max} - f^R_{i,\min}} \right)^2}{\sum_{i=1}^{n} w_i} \right]^{\frac{1}{2}} = \left[\frac{1.0 \cdot ABS\left(\frac{3-2}{3-0} \right)^2}{1.0} \right]^{\frac{1}{2}} = 0.33$$

In the above calculation, a quality score of 0.33 is calculated for the k^{th} retrieved case as it has accounted for two of the three confounding factors. Alternatively, if none, one, or all of the three confounding factors are treated, a quality score of 1.0, 0.66, or 0.0 can be computed respectively. Clearly, the smaller the quality score, the more valid the case and its corresponding solutions.

3.5 Relevance Score

Not all cases retrieved from the ISECR case base, based on querying for the same countermeasure and location types, have the same location characteristics as the design case. For example, if the design case deals with an urban signalized intersection and a total entering traffic volume of 10000 AADT, not all of the retrieved cases have the same site features as the design case. Hence, it would be reasonable to assign more weight to a case and its results if it has more similar site features compared to the design case.

Relevance score determined for each case is influenced by the degree of similarity in the location characteristics between the design and retrieved cases, as provided by the ISECR user and those already stored in the case base respectively. The location characteristics considered, as mentioned previously, are different for each of the eight location types (see Appendix B for details).

3.5.1 Calculation of Relevance Scores

In this thesis, relevance score for each retrieved case is determined by utilizing the Equation 3.1. To calculate the relevance score using Equation 3.1, the values of $f_{i,\max}^{R}$ and $f_{i,\min}^{R}$ assigned to each feature of a case are based on the specific feature encountered. For example, suppose that the encountered feature is traffic volume, values of 11.0 and 1.0 are assigned to $f_{i,\max}^{R}$ and $f_{i,\min}^{R}$ to represent the maximum traffic volume of 80000 and more AADT and the minimum traffic volume of 0-4999 AADT respectively. As for the values of f_i^I and $f_{i,k}^R$, they are based on the specific features of the design and retrieved cases respectively (see Appendix B for details). Each feature is weighted equally in Equation 3.1, i.e., the value of w_i is dependent on the number of features considered. For instance, a value of 0.5 is assigned if two features are considered. Similarly, a value of 0.33, 0.25, or 0.20 is assigned to w_i if three, four, or five, features are considered respectively.

To demonstrate how Equation 3.1 can be used to determine the relevance score for a retrieved case, consider the example below where the design case deals with an urban signalized intersection with a total entering traffic volume of 10000 AADT, while the retrieved case concerns with a suburban signalized intersection with a total entering traffic volume of 5000 AADT. These features are represented in ISECR as follows:

Table 3.2. Example: Representation of the Location Characteristics

	Feature Values			
	Maximum	**Minimum**	**Design Case**	**Retrieved Case**
Area type	4	1	1 (urban)	2 (suburban)
Total Entering traffic volume (AADT)	11	1	3 (10000-14999)	2 (5000-9999)

The relevance score for the retrieved case can be determined by using Equation 3.1 as:

$$D_{k(\mathrm{Re}\,levance)} = \left[\frac{\sum_{i=1}^{n} w_i \cdot ABS\left(\frac{f_i^I - f_{i,k}^R}{f_{i,\max}^R - f_{i,\min}^R} \right)^2}{\sum_{i=1}^{n} w_i} \right]^{\frac{1}{2}} = \left[\frac{0.5 \cdot ABS\left(\frac{1-2}{4-1} \right)^2 + 0.5 \cdot ABS\left(\frac{3-2}{11-1} \right)^2}{0.5 + 0.5} \right]^{\frac{1}{2}}$$

$$D_{k(\mathrm{Re}\,levance)} = 0.246$$

A lower relevance score signifies that the location characteristics are more similar between the design and retrieved cases, and vice versa.

3.6 Calculating Case Distance in ISECR

Case distance is evaluated to determine the relative importance of a case compared to others. When computing case distance, D_k, for each of the retrieved cases, the ISECR user can decide to include either the quality score alone or a combination of the quality and relevance scores. If the user decides to exclude the relevance score and only uses the quality score, D_k is equivalent to the quality score computed for the retrieved case. However, if the user decides to encompass both the quality and relevance scores in determining D_k, both scores are weighted equally as in the expression below:

$$D_k = \frac{1}{2} \cdot \left(D_{k(Quality)} + D_{k(\mathrm{Re}\,levance)} \right) \qquad \textit{(Equation 3-2)}$$

where D_k = Case distance of the k^{th} retrieved case

For example, with the quality and relevance scores (0.33 and 0.246) computed for the two examples in Sections 3.4.4 and 3.5.1 respectively, case distance for the retrieved case can be determined in the following two ways:

$$D_k = D_{k(Quality)} = 0.33$$

$$D_k = \frac{1}{2} \cdot \left(D_{k(Quality)} + D_{k(\mathrm{Re}levance)}\right) = \frac{1}{2} \cdot (0.33 + 0.246) = 0.288$$

The first result indicates that the ISECR user has decided to exclude the relevance score in determining the case distance for the retrieved case. If this is the case, case distance is exactly the same as the quality score determined, i.e., in this example, 0.33. Conversely, the second result indicates that both the quality and relevance scores are utilized in determining D_k. In this instance, the two scores are weighted equally and determined to be 0.288.

Hence, a lower case distance is assigned to a case if it has treated more confounding factors and/or if its location characteristics are more similar to the design case.

3.7 Solution Construction in ISECR

In ISECR, solutions to the design problem can be constructed utilizing either the nearest neighbour or collaborative approaches (see Section 2.2.5 for details).

3.7.1 Nearest Neighbour Approach

The nearest neighbour approach utilizes the most relevant case (lowest case distance) and its result is used as the solution to the design case. The ISECR user is presented with a list of all matched cases ranked in an ascending order according to the calculated case distances. The user can then select the most appropriate case and adapt the case solutions (CRFs), if necessary, to account for the differences between the design and retrieved cases.

3.7.2 Collaborative Approach

The collaborative CBR approach, on the other hand, utilizes more than one retrieved case, combines their results, and then employs the combined result as the solution to the current problem. With this approach, the *k*-nearest neighbour retrieval is used, where *k* is the number of cases to be retrieved and utilized in solution combination. There is little evidence in the literature indicating what the optimum *k* value should be to produce the best results. Not only that, the *k* values found in the literature vary greatly from one application to the next. For example, Gonzalez and Laureano-Ortiz (1992) combined outcomes from the three most relevant cases (if available) while Yeh (1997) utilized the results stored in all of the retrieved cases.

The approach employed in ISECR uses all of the retrieved cases to derive the new solution (the effectiveness of the proposed countermeasure and its uncertainty). Although this approach may include results from cases that have lower case distances, it should be realized that utilizing solutions stored in the most similar cases may not necessarily provide the best solution to the new problem (Yeh, 1997). Furthermore, this report intends to provide both the non-weighted average and weighted average solutions (CRFs) based on the results obtained from all of the retrieved cases. The non-weighted average is simply the arithmetic mean of the retrieved CRFs, while the weighted result is determined by weighting the retrieved CRFs with their case distances. Equations 3.3 and 3.4 are used in ISECR to determine the non-weighted and weighted CRFs respectively:

$$CRF_{Non-Weighted} = \frac{\sum_{k=1}^{N} CRF_k}{N} \qquad \textit{(Equation 3.3)}$$

$$CRF_{Weighted} = \frac{\sum_{k=1}^{N} (1 - D_k) \cdot CRF_k}{\sum_{k=1}^{N} (1 - D_k)} \qquad \textit{(Equation 3.4)}$$

where $CRF_{Non-Weighted}$ = Non-weighted solution (CRF)

$CRF_{Weighted}$ = Weighted solution (CRF)

N = Number of retrieved cases

D_k = Case distance of the k^{th} retrieved case

CRF_k = CRF of the k^{th} retrieved case

The accuracy of the above results can be estimated with the standard deviations as:

$$\sigma_{CRF-Non-Weighted} = \sqrt{\frac{\sum_{k=1}^{N}(CRF_k - CRF)^2}{N-1}} \qquad \textit{(Equation 3.5)}$$

$$\sigma_{CRF-Weighted} = \sqrt{\frac{\sum_{k=1}^{N}(1-D_k)\cdot(CRF_k - CRF)^2}{\sum_{k=1}^{N}(1-D_k)\cdot(N-1)}} \qquad \textit{(Equation 3.6)}$$

where $\sigma_{CRF-Non-Weighted}$ = Standard deviation of the non-weighted solution

$\sigma_{CRF-Weighted}$ = Standard deviation of the weighted solution

The provision of the weighed results attempts to account for either the quality or both the quality and relevance of a case, i.e., solutions obtained from cases with lower case distances are weighted more and vice versa. This is demonstrated in the following example as shown in Table 3.3:

Table 3.3. Example: Determining Non-Weighted and Weighted CRFs

Ranking of the retrieved case	Case Distance, D_k	$(1-D_k)$	Case Result, CRF_k (%)	Weighted Case Result (%)
1	0.25	0.75	23	17.25
2	0.5	0.5	25	12.5
3	0.75	0.25	33	8.25
4	1.0	0.0	40	0
Sum	**2.5**	**1.5**	**121.0**	**38.0**

The above table demonstrates that cases with lower case distances are weighted considerably higher than those with higher case distances. The non-weighted and weighted CRFs can be determined by using Equations 3.3 and 3.4 respectively, as shown below:

$$CRF_{Non-Weighted} = \frac{\sum_{k=1}^{N} CRF_k}{N} = \frac{121.0}{4.0} = 30.25\%$$

$$CRF_{Weighted} = \frac{\sum_{k=1}^{N} (1 - D_k) \cdot CRF_k}{\sum_{k=1}^{N} (1 - D_k)} = \frac{38.0}{1.5} = 25.33\%$$

Hence, the non-weighted and weighted CRFs are 30.25% and 25.33% respectively. In this example, the weighted CRF is less than the non-weighted CRF. This is expected as the weighted CRFs reflect results that have been accounted for some or all of the three confounding factors, as mentioned previously. In some instances, ISECR provides weighted CRFs that are slightly larger than the non-weighted results. This arises mainly due to the insufficient data available and/or the small number of evaluation studies available for the specific safety improvement implemented at the specific location.

Continuing with the example shown in Table 3.3, the standard deviations for the non-weighted and weighted CRFs can be estimated with Equations 3.5 and 3.6 respectively as:

$$\sigma_{CRF-Non-Weighted} = \sqrt{\frac{\sum_{k=1}^{N} (CRF_k - CRF)^2}{N-1}} = \sqrt{\frac{182.75}{3}} = 7.81\%$$

$$\sigma_{CRF-Weighted} = \sqrt{\frac{\sum_{k=1}^{N}(1-D_k)\cdot(CRF_k - CRF)^2}{\sum_{k=1}^{N}(1-D_k)\cdot(N-1)}} = \sqrt{\frac{18.83}{4.5}} = 2.05\%$$

Thus, the standard deviations for the non-weighted and weighted CRFs are 7.81% and 2.05% respectively. Clearly, the non-weighted standard deviation is significantly greater when compared to the weighted standard deviation.

4.0 DESCRIPTION OF ISECR

4.1 Main Menu

ISECR is designed as an user-friendly intelligent database that can facilitate the daily use of practitioners in the highway safety engineering industry. A prototype of ISECR is currently implemented in a personal computer using Microsoft Access 97 in a Microsoft Windows 95/98 environment. Currently, the database consists of 450 documents.

In addition to predicting the effectiveness of safety improvements and their reliability, ISECR is designed to accept queries, filter and display information based on the users' specifications, and permit the entry of new documents. This chapter provides a description of the features available in ISECR. Specifically, the first four of the following five ISECR *Main Menu* options are discussed (see Figure 4.1):

1. View all documents,
2. Search documents,
3. Predict CR factors,
4. Input new documents, and
5. Exit (ISECR).

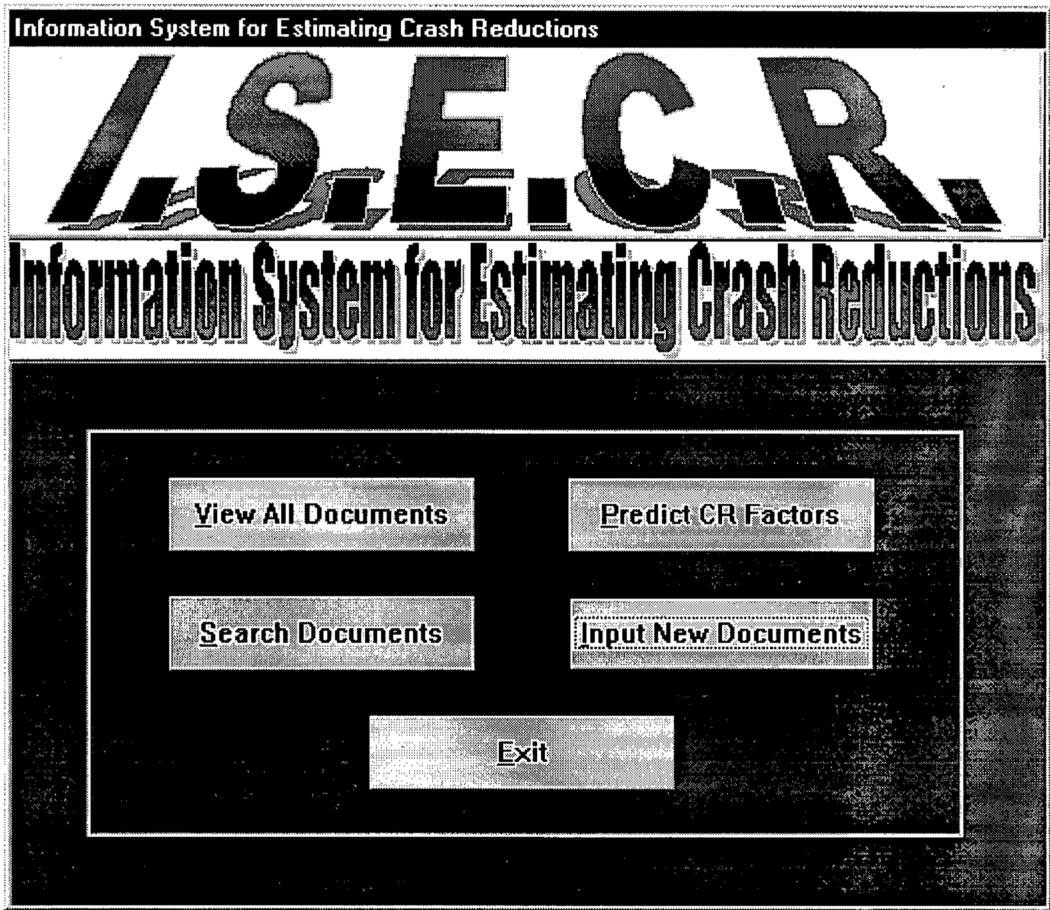


Figure 4.1. ISECR Window: Main Menu

4.2 View All Documents

With the *View All Documents* option, the user can list all evaluation studies stored in the ISECR database. As illustrated in Figure 4.2, ISECR allows the user to list and sort the available studies based on:

1. Author name (including all authors and co-authors),
2. Study source,

3. Study title, or
4. Publication date.

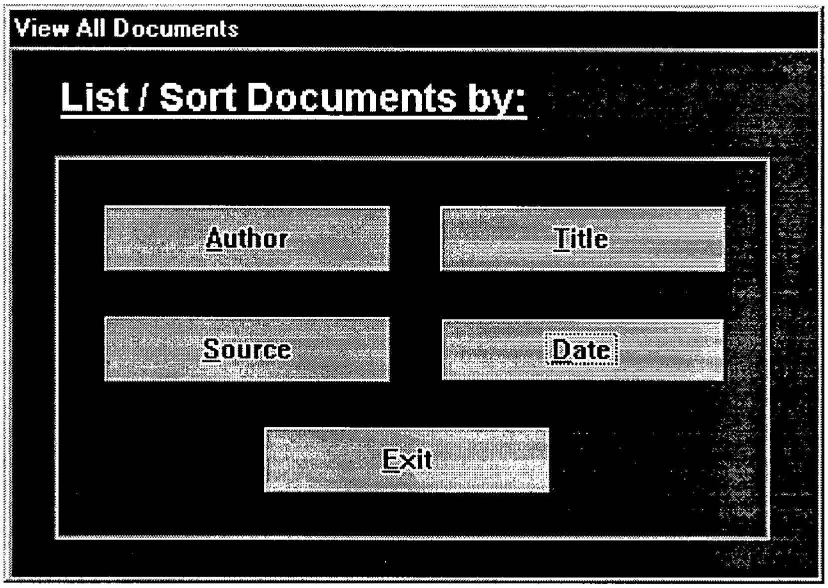


Figure 4.2. ISECR Window: View All Documents

Once the sorting option is selected, ISECR then presents a summary of the documents including the following information (see Figure 4.3):

1. Authors' names,
2. Study title,
3. Study source,
4. Publication date,
5. Availability of the one page summary for the document, and
6. Availability of the document.

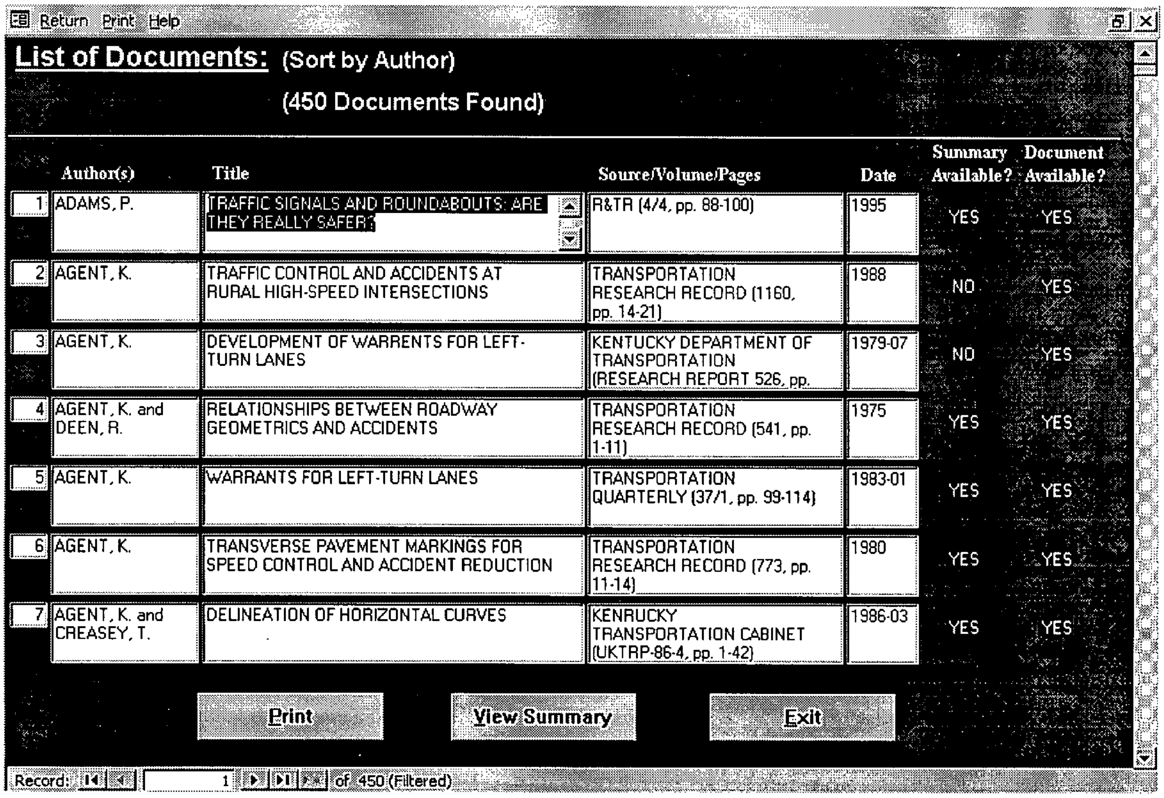


Figure 4.3. ISECR Window: Summary of Documents

If the one page summary of the document is available, the user can click on the *View Summary* button to view the summary of the document. Figure 4.4 shows an example of the document summary. The available summaries in the ISECR database are provided by G.D. Hamilton & Associates Consulting LTD.

Author	**Title**	**Reference**	**Year**	**County**
Adams P. File: adams.dot	**Traffic signals and roundabouts: Are they really safer?**	R&TR Vol. 4 No. 4	1995	Aus

Location Sydney, Australia.

Level System wide. 11 traffic signal and 13 roundabouts sites studied .

Methodology Two years before and after accident data. A control was introduced using the entire Local Government Area.

Shortcomings No information on roundabouts and intersections characteristics

Road and vehicle characteristics NA

Accident cause and pattern Three classes of accident severity:

- PDO
- Non-admitted injury
- admitted injury and fatalities

Safety countermeasure or Strategy Traffic signals and roundabouts

Effectiveness

	Mean of % change	Percentage change	Adjusted percentage changes (control section)
Traffic signal at intersection	-31.2 (stdev 63.04)	-41.7	-35.1
Roundabouts	-61.7 (stdev 39.8)	-77.7	-71.1

See Table 3 for detail with accident categories

Miscellaneous NA

NA = Non Applicable or Non Available.
Ns = not significant.

Figure 4.4. ISECR Window: An Example of One Page Summary

4.3 Search Documents

This feature of the software provides the user a search capability on documents by entering a set of query parameters. The following query parameters are included (see Figure 4.5):

1. Author name,
2. Publication year,
3. Country,
4. Countermeasure, and
5. Location.

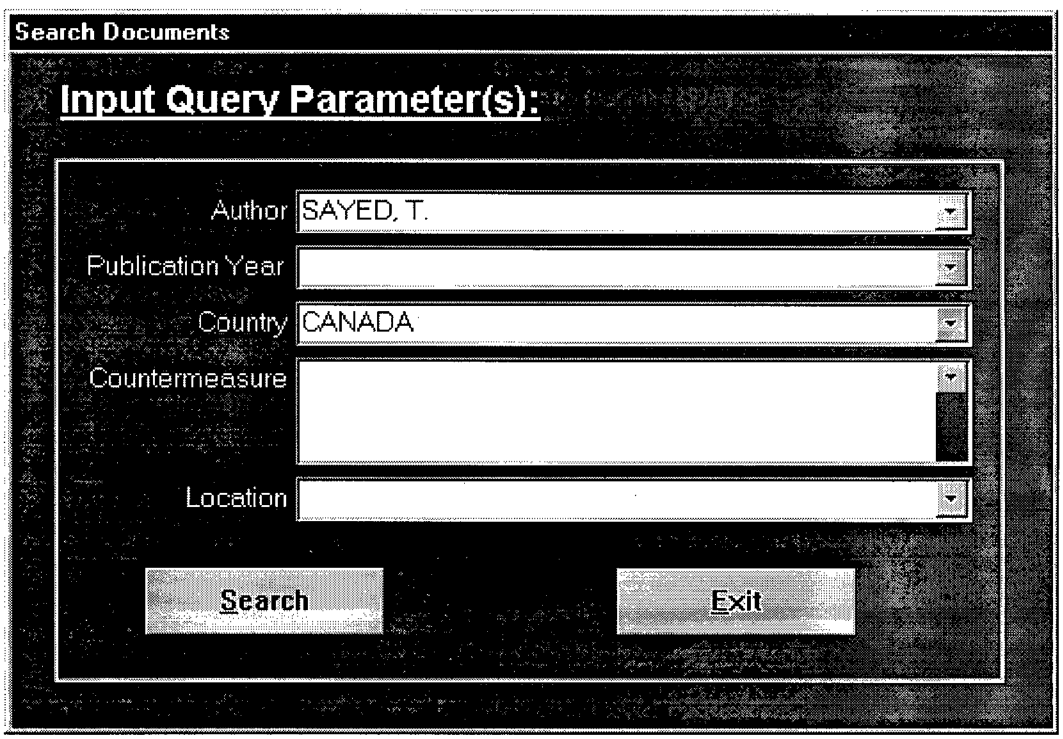


Figure 4.5. ISECR Window: Selecting Query Parameters

Once the query parameters are selected, the database is searched and a summary of the matched documents is presented similarly as shown in Figure 4.3.

4.4 Predict CR Factors (CRFs)

This option allows the user to input the specifications of the new problem and to predict the effectiveness of a specific safety improvement. Utilizing a CBR approach, as explained in Chapters 2 and 3, past records reporting CRFs can be retrieved and analyzed to generate results for the current problem. Once the *Predict CR Factors* option is chosen, the user is first presented a list of sixteen countermeasure categories as illustrated in Figure 4.6.

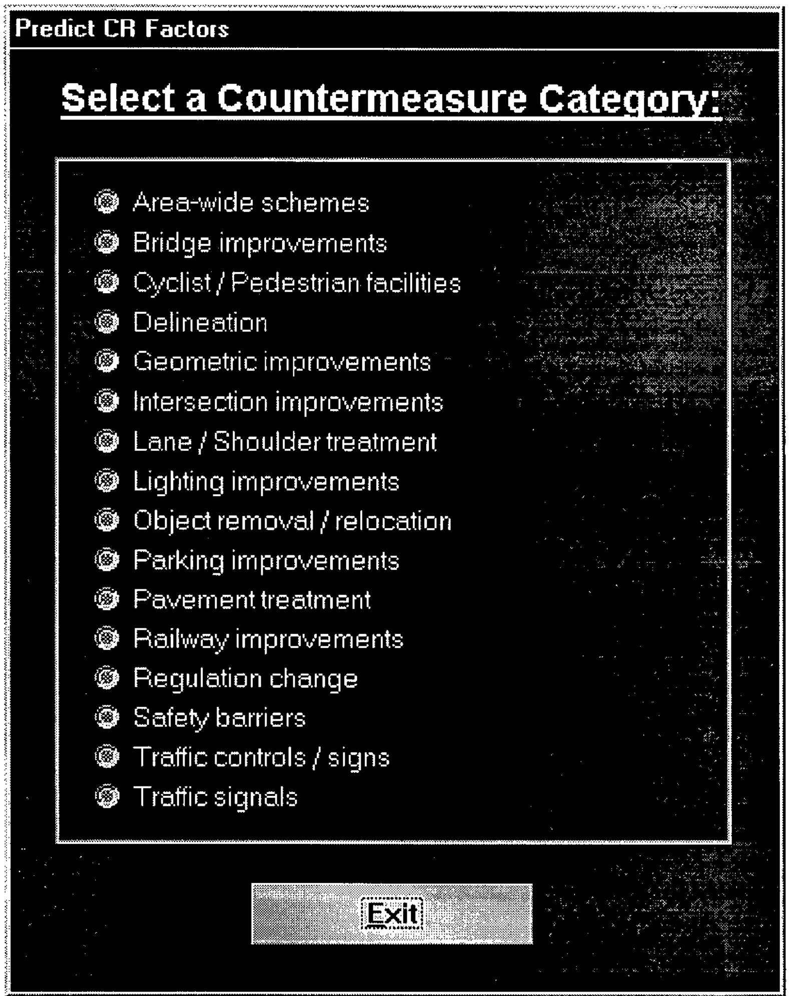


Figure 4.6. ISECR Window: Countermeasure Categories

By clicking on any one of the sixteen categories, another window consists of different countermeasures (belonging to the selected countermeasure category) is opened. For example, Figure 4.7 is a result of selecting *Delineation* as the countermeasure category. Appendix C provides the details when other countermeasure categories are selected.

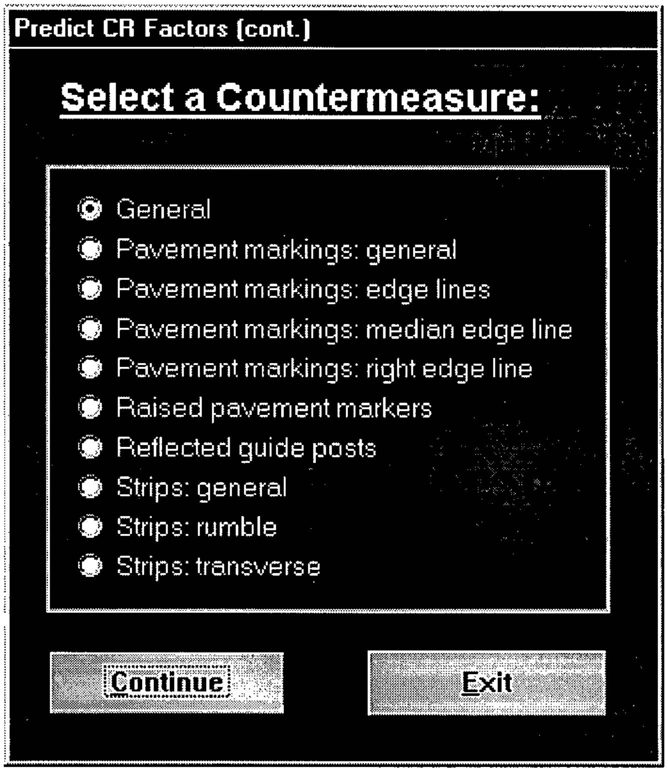


Figure 4.7. ISECR Window: Countermeasure Types for Delineation

Once the countermeasure type is selected, the user is required to click on the *Continue* button before the location type window can be opened. This window allows the user to specify where the proposed countermeasure is to be implemented. The location type window, as shown in Figure 4.8, consists of eight different location types.

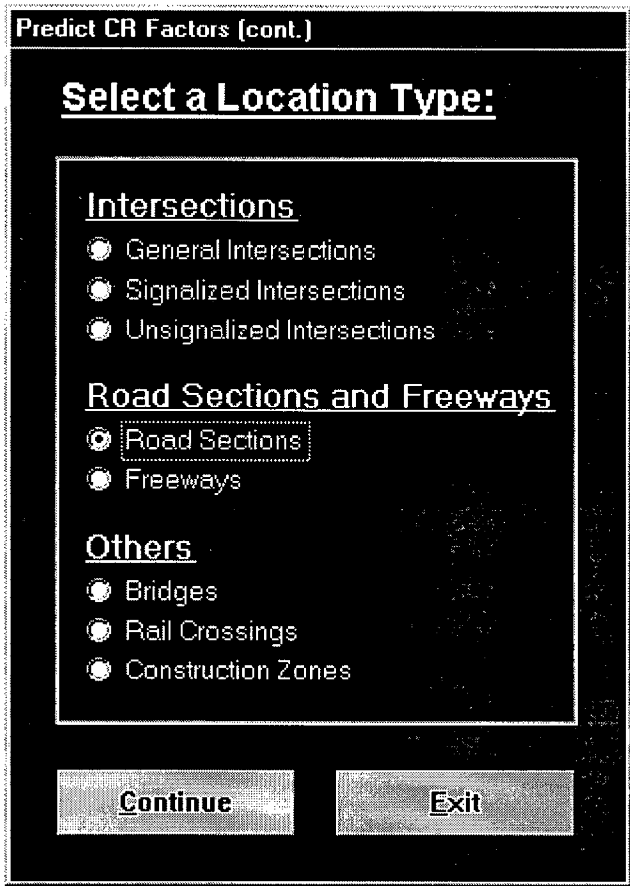


Figure 4.8. ISECR Window: Location Types

After the location type is specified, ISECR searches its case base for cases dealing with the same countermeasure and location types. A case is only retrieved if a perfect match of the two query parameters is realized. If no matched case is found with the specified query parameters, the user will be notified, as illustrated in Figure 4.9:

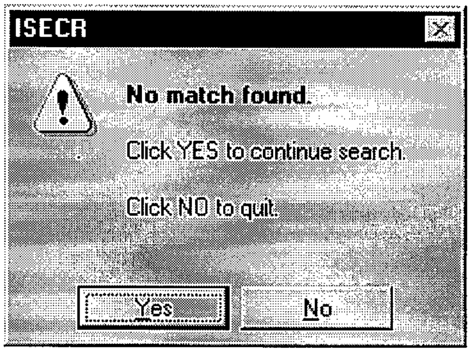


Figure 4.9. ISECR Window: No Match Found

On the other hand, if cases are found and retrieved, the user will then be prompted with a dialog box to enter the location characteristics which would be used later in the calculation of case distances. This is shown in Figure 4.10. Once the *Yes* button is clicked, the user is presented with an input form, as shown in Figure 4.11 for signalized intersections, to enter location characteristics (see Appendix D for the input forms used for other location types). In ISECR, the implementation level for the design case is assumed to be at an isolated location. This assumption is made if the user decides to use location characteristics in the analysis. Once the user finishes entering location characteristics and clicks on the *Continue* button, he/she is warned by ISECR to use quality criterion only if few location characteristics were entered. As shown in Figure 4.12, the user can then decide if case distances will be calculated based on the quality scores alone or based on both the quality and relevance scores.

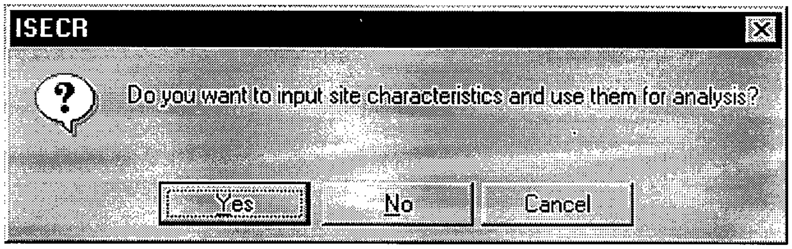


Figure 4.10. ISECR Window: Selecting to Enter Location Characteristics

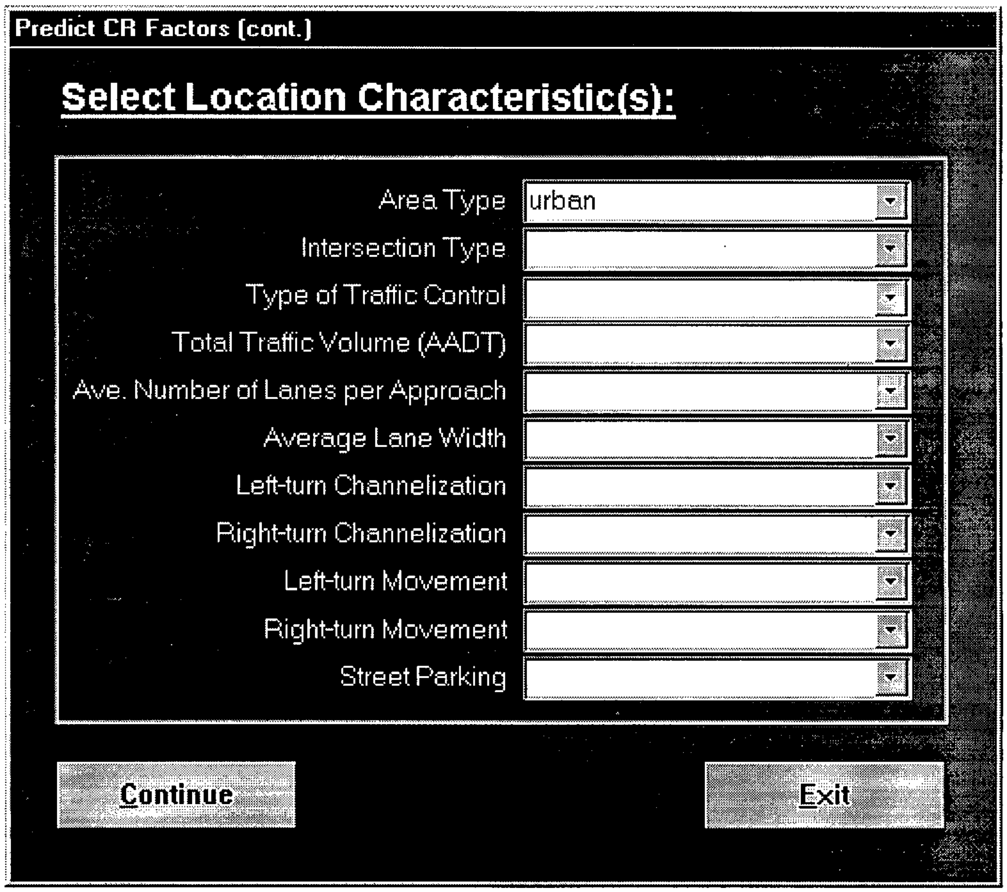


Figure 4.11. ISECR Window: Input Location Characteristics for Signalized Intersections

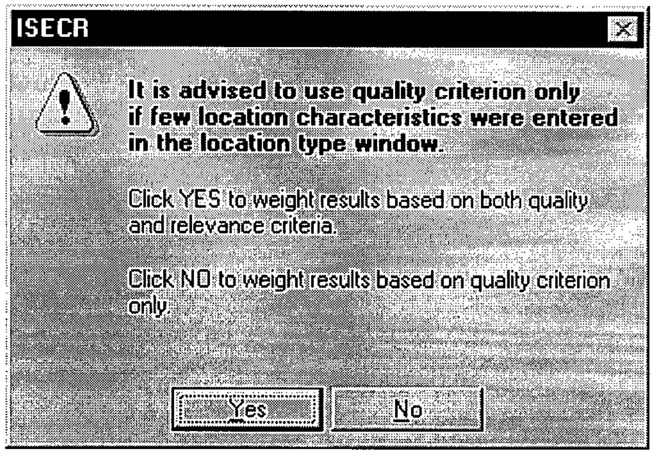


Figure 4.12. ISECR Window: Selecting the Criteria to Calculate Case Distances

However, if *No* is clicked in the ISECR window (Figure 4.10) to indicate that no location characteristics will be entered, case distances will only be calculated based on considering the quality scores alone. Whether the user decides to use only the quality scores or both the quality and relevance scores in determining the case distances, ISECR then presents a summary of ranked evaluation studies, which are essentially the retrieved cases that match the query parameters. This summary is displayed in Figure 4.13.

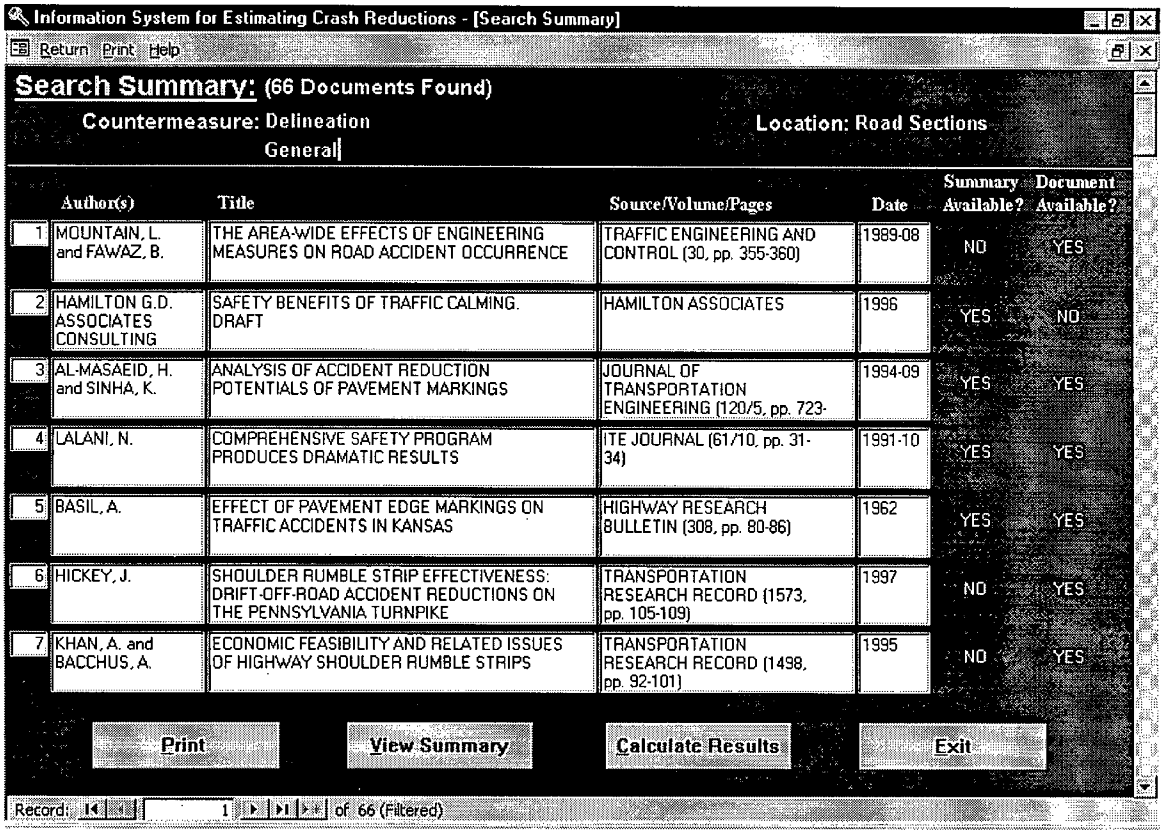


Figure 4.13. ISECR Window: Summary of Ranked Documents

Since some evaluation studies may consist of more than one case that matched the query parameters, the ranking of the studies is achieved by assigning each study a case distance only when the calculated case distance for that retrieved case is the lowest among all of its retrieved cases. The studies are then ranked according to the case distance assigned, as shown in the following example:

Table 4.1. Example: Ranking of Evaluation Studies

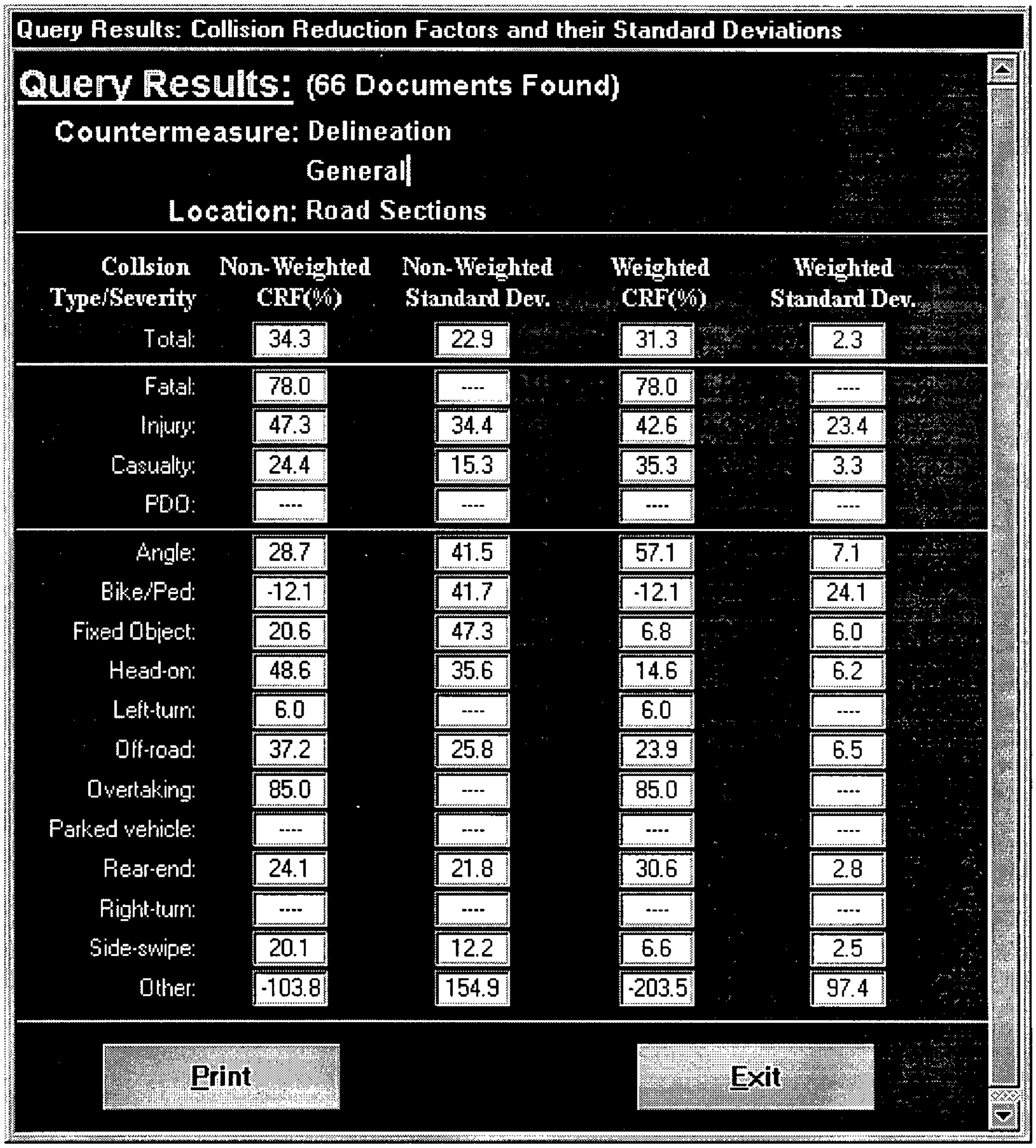
Study (Study Id)	Retrieved Case (Case Id)	Case Distance	Case Distance assigned to the Study?
1	***1***	***0.1***	***Yes***
2	***3***	***0.125***	***Yes***
1	2	0.25	No
3	***6***	***0.25***	***Yes***
4	***7***	***0.3***	***Yes***
2	4	0.3	No
2	5	0.3	No
4	8	0.3	No
4	9	0.3	No
5	***10***	***0.35***	***Yes***

Hence, with the above example, the following five studies are listed in the ISECR summary window (see Figure 4.13):

Table 4.2. Example: Final Output to the Summary Window

Ranking	Study (Study Id)	Retrieved Case (Case Id)	Case Distance
1	1	1	0.1
2	2	3	0.125
3	3	6	0.25
4	4	7	0.3
5	5	10	0.35

Finally, to determine the countermeasure effectiveness, the user can click on the *Calculate Results* button, as shown in Figure 4.13. As mentioned in Chapter 3, ISECR utilizes all of the retrieved cases to derive the solutions (CRFs and the associated standard deviations) required for the current problem. Thus, with the same example as the one shown in Table 4.1, a total of ten cases from five evaluation studies would be used. The predicted CRFs and their standard deviations for total, and various collision severity and types are presented in an ISECR window as shown in Figure 4.14. The non-weighted and weighted CRFs are determined by using Equations 3.3 and 3.4 respectively. As for the non-weighted and weighted standard deviations, Equations 3.5 and 3.6 are used respectively.


Query Results: Collision Reduction Factors and their Standard Deviations

Query Results: (66 Documents Found)

Countermeasure: Delineation

General

Location: Road Sections

Collision Type/Severity	Non-Weighted CRF(%)	Non-Weighted Standard Dev.	Weighted CRF(%)	Weighted Standard Dev.
Total:	34.3	22.9	31.3	2.3
Fatal:	78.0	----	78.0	----
Injury:	47.3	34.4	42.6	23.4
Casualty:	24.4	15.3	35.3	3.3
PDO:	----	----	----	----
Angle:	28.7	41.5	57.1	7.1
Bike/Ped:	-12.1	41.7	-12.1	24.1
Fixed Object:	20.6	47.3	6.8	6.0
Head-on:	48.6	35.6	14.6	6.2
Left-turn:	6.0	----	6.0	----
Off-road:	37.2	25.8	23.9	6.5
Overtaking:	85.0	----	85.0	----
Parked vehicle:	----	----	----	----
Rear-end:	24.1	21.8	30.6	2.8
Right-turn:	----	----	----	----
Side-swipe:	20.1	12.2	6.6	2.5
Other:	-103.8	154.9	-203.5	97.4

Print Exit

Figure 4.14. ISECR Window: Predicted CRFs and their Standard Deviations

4.5 Input New Documents

The *Input New Documents* option allows new documents to be added to the ISECR database. However, in order to assure the quality and consistency of data entry, input of new documents in ISECR is limited to the agency responsible for the database maintenance. As far as the types of information entered for each document and how they are represented in the database, Section 3.2 should be referred.

5.0 VALIDATION OF ISECR

5.1 Introduction

Validation is an essential step to confirm the success of any computer software. In order to validate the results produced by ISECR, this chapter is devoted to compare the CRFs predicted by ISECR and those extracted from the literature for 15 safety improvements (see Table 5.1 for details). These improvements were selected randomly from the 116 countermeasures considered in ISECR (see Appendix A for details).

5.2 ISECR Results in the Context of Published Studies

Pertinent portions of four previously published sources (Tamburri and Smith, 1971, Creasey and Agent, 1985, McFarland et al., 1978, Terry and Watson, 1982) that have compiled and tabulated a summary of CRFs for various countermeasures are summarized in Table 5.1. Table 5.1 also provides both the non-weighted and weighted CRFs produced by ISECR for the chosen 15 countermeasures.

Upon inspection of Table 5.1, it is evident that the scope of each study varies greatly from one to another. The CRFs presented in Table 5.1, while representative of the four published studies, should not be considered to be complete as there are additional collision data presented in these reports which are not relevant for the purpose of this thesis and thus, excluded from Table 5.1.

Table 5.1. Summary of CRFs: ISECR Results vs. Published Results

	CRF (Total Collisions)					
Countermeasure Type	**Tamburri & Smith 1971**	**Creasey & Agent 1985**	**McFarland et al. 1978**	**Terry & Watson 1982**	**ISECR Non-Weighted**	**ISECR Weighted**
Bridge improvements						
Widen bridges		30-65		65	63.5	
Geometric improvements						
Horizontal alignment		20-40	40-88	41	25.6	23.8
Vertical alignment		15-54			30.3	29.1
Hor & vert alignment		50-52		20-21	40.4	32.7
Sight distance		20-31		31	33.5	
Intersection improvements						
Left-turn chan.		15			38.2	31.9
Lighting improvements						
Install at intersections	75[b]	75[b]	50[c]		43.8, 43.8[c]	36.6, 40.3[c]
Pavement treatment						
Pavement grooving	75[a]	10-48			58.3	
Resurfacing		12-42	12-44		22.4	19.0
Skid reduction		13-50	21		31.4	14.7
Railway improvements						
Flashing beacons		70-94			64.5	60.3
Traffic controls/signals						
4-way stops	70	68-70		73	62.8	44.1
Traffic signals						
Flashing beacons			37		54.0	
red-yellow	50	34		26	34.0	
New signals	15	15-80	6-29	20	25.6	21.0

[a]Wet pavement collisions only

[b]Night time collisions only

[c]Rural area type

By examining Table 5.1, it is evident that there is a general agreement between the ISECR results and the published results. In some instances, the ISECR CRFs are considerably higher than those reported by the published literature. For example, while McFarland et al. (1978) reported a CRF of 37% for implementing flashing beacons, ISECR predicted a CRF of 54%. The variation may be due to the difference in the sample size of studies and the approach undertaken by each study in determining the CRFs. For instance, by examining the ISECR outputs, it is noted that a small sample size of studies is used in the determination of the CRF, i.e., only four studies are retrieved from the ISECR database. Not only that, of the four studies retrieved, three (Mayer, 1971, Cribbins and Walton, 1970, Wilson, 1967) of them are outdated and analyzed their collision data by using the simple before and after approach. Although the studies are outdated, the ISECR result is an indication of how effective the countermeasure is, without correcting for the confounding factors. Further examination of the ISECR outputs for other countermeasures, it is also noted that the ISECR predictions are more valid and comparable to the literature results when there is a larger sample size available for the specific countermeasure. Based on the countermeasures examined in Table 5.1, it is apparent that more work is required to enter new evaluation studies into the ISECR database to further increase the reliability of the ISECR results for some countermeasures. Nevertheless, for the majority of the countermeasures examined above, ISECR does provide valid results.

6.0 EXPECTED COLLISION REDUCTION

6.1 Introduction

The safety benefits of implementing a safety improvement can be represented by the expected reduction in collision frequency, which is normally calculated by the product of CRF and the expected number of collisions (denoted by N).

This thesis has proposed the use of a CBR approach, one of the recent developments in problem-solving paradigms in artificial intelligence, to assist in developing an intelligent database (ISECR) that is capable of assessing the range and reliability of the predicted countermeasure effectiveness. With ISECR, CRFs are determined on a project-level basis where the variability of geometric configuration among different locations is considered. Moreover, ISECR addresses the uncertainty issues related to the effectiveness of the proposed safety improvement. Specifically, non-weighted and weighted CRFs and their standard deviations can be determined with ISECR as shown previously.

The expected number of collisions, N, at a given location can be evaluated by procedures such as the multivariate approach, i.e., the GLIM (generalized linear modeling) approach, and the EB (Empirical Bayes) approach. As discussed in Chapter Two, these techniques can be used to readily provide more accurate site-specific safety estimates compared to the conventional approaches. The outcomes from these techniques are drawn upon in the next section, i.e., the expected value and standard deviation of N.

The next section outlines the procedures involved in evaluating the collision reduction and its uncertainty expected from implementing a safety improvement. Specifically, the moment approach (Benjamin and Cornell, 1970, and Ang and Tang, 1984) is utilized to combine CRFs and expected number of collisions and their uncertainties.

6.2 Collision Reduction and Its Uncertainty

With the availability of the expected values and standard deviations of both CRF and N, the expected reduction in collision frequency, Z, and its variance can now be calculated. Assuming that CRF and N are independent of each other, i.e., no correlation between the two variables, the expected value of Z can be determined by the following equation:

$$E(Z) = N \times CRF \qquad \textit{(Equation 6.1)}$$

where $E(Z)$ = Expected reduction in collision frequency in the after period

N = Expected number of collisions

CRF = Collision reduction factor

The accuracy of Z is represented by the variance which can be calculated as:

$$Var(Z) = N^2(\sigma_{CRF})^2 + CRF^2(\sigma_N)^2 + (\sigma_{CRF} \times \sigma_N)^2 \qquad \textit{(Equation 6.2)}$$

where $Var(Z)$ = Variance of Z

σ_{CRF} = Standard deviation of CRF

σ_N = Standard deviation of N

The application of the above procedures is illustrated with an example in Chapter Eight.

7.0 ECONOMIC ANALYSIS

7.1 Introduction

There are a number of methods that can be used to evaluate the economic feasibility of implementing a safety improvement. One of the most frequently used evaluation measures is the benefit-cost ratio (BCR). The BCR is a measure of the amount of dollar return expected with every dollar spent on a safety improvement. The expected benefits from implementing a countermeasure can be evaluated by the savings anticipated from the reduction in collisions. With the expected value and variance of Z determined in Chapter Six, the uncertainty issues associated with BCR can now be addressed. Specifically, the probability of a proposed countermeasure achieving a preset BCR can now be determined.

7.2 Benefit-Cost Ratio (BCR) and Its Uncertainty

The expected value of BCR for implementing a countermeasure can be determined by the following equation:

$$E(BCR) = \frac{(Col.Cost) \times E(Z) \times (P/A,i,t)}{Cost_{implementation}} = k \times E(Z) \qquad \textit{(Equation 7.1)}$$

where $E(BCR)$ = Expected value of BCR

$E(Z)$ = Expected reduction in collision frequency in the after period

$Col.Cost$ = Average collision cost

t = Payback period (year)

i = Discount rate

$(P/A,i,t)$ = Present worth factor, given the payback period and discount rate

$$k = \frac{(Col.Cost) \times (P/A,i,t)}{Cost_{implementation}} \quad \textit{(Equation 7.2)}$$

Assuming that the only random variable in Equation 7.1 is Z, i.e., k is a constant variable, the variance of BCR can then be expressed by Equation 7.3:

$$Var(BCR) = k^2 \times Var(Z) \quad \textit{(Equation 7.3)}$$

where $Var(BCR)$ = Variance of BCR

$Var(Z)$ = Variance of Z, as defined by Equation 6.2.

7.3 Probability Density Function of BCR

With the expected value and variance of BCR, the probability density function of BCR can now be established. This thesis utilizes a Gamma distribution to model BCR, as this distribution is suitable for modeling continuous random variable and providing wide variety of shapes. Furthermore, this distribution is also limited to positive values and skewed to the right, which is appropriate to model BCR. The Gamma distribution parameters, α and β, and the probability density function of BCR can be calculated by the following equations:

$$\beta = \frac{E(BCR)}{Var(BCR)} \qquad \textit{(Equation 7.4)}$$

$$\alpha = E(BCR) \times \beta \qquad \textit{(Equation 7.5)}$$

$$f(BCR;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(BCR)^{\alpha-1}e^{(-\beta\times(BCR))} \qquad \textit{(Equation 7.6)}$$

where α and β = Gamma distribution parameters

$f(BCR;\alpha,\beta)$ = Probability density function of BCR

Once the above probability density function is defined, the cumulative distribution function for BCR can be formulated as follows:

$$F(BCR;\alpha,\beta) = \int_0^1 f(BCR;\alpha,\beta)d(BCR) \qquad \textit{(Equation 7.7)}$$

where $F(BCR;\alpha,\beta)$ = Cumulative distribution function of BCR

By plotting the cumulative distribution function of BCR, the probability of a countermeasure achieving a specific BCR upon its implementation can be determined. This is illustrated with an example in the next chapter.

8.0 APPLICATIONS

8.1 Problem Definition

The procedures outlined in the previous chapters can be used to evaluate the cost-effectiveness of road safety improvements. For example, assume that a traffic safety engineer is interested in determining the probability of achieving a 2:1 return, i.e., a benefit-cost ratio (BCR) of 2.0, by improving public lighting along an arterial street (or known as road sections in the location type feature used in this research) that has the following characteristics:

- Area type: suburban
- Implementation level: isolated location (this is automatically assumed by ISECR when the user decides to include the relevance criterion in the calculation of case distances)
- Total traffic volume: 25000 AADT

With the above problem specifications, the remaining sections in this chapter intend to:

1. Determine the effectiveness of the proposed countermeasure, i.e., CRFs and the standard deviations.
2. Calculate the reduction in collision anticipated once the countermeasure is implemented by utilizing the moment approach, as discussed in Chapter Six.
3. Compute the BCR and its variance for the proposed countermeasure and plot its cumulative distribution function to assess the probability for the proposed countermeasure achieving a BCR of 2.0.

8.2 Countermeasure Effectiveness

The effectiveness of the proposed countermeasure is evaluated by utilizing cases currently available in the ISECR case base. Relevant cases are retrieved only if the same countermeasure and location types match between the design case and cases stored in the ISECR case base. Querying the ISECR case base with general lighting improvements and road sections as the countermeasure and location types respectively, a total of 53 cases from 38 documents are retrieved. Table 8.1 provides a list of the retrieved cases.

To determine the case distance for each of the retrieved cases, the quality and relevance scores are considered. Tables 8.1 and 8.2 summarize the quality (treatment of the confounding factors) and relevance (location characteristics) information for each case, as extracted from Subcases 2 and 5 from the ISECR case base respectively. With the information available, both the quality and relevance scores can be determined by using Equation 3.1, as shown below for Case Id 296:

$$D_{k(Quality)} = \left[\frac{\sum_{i=1}^{n} w_i \cdot ABS\left(\frac{f_i^I - f_{i,k}^R}{f_{i,\max}^R - f_{i,\min}^R} \right)^2}{\sum_{i=1}^{n} w_i} \right]^{\frac{1}{2}} = \left[\frac{1.0 \cdot ABS\left(\frac{3-1}{3-0} \right)^2}{1.0} \right]^{\frac{1}{2}} = 0.667$$

$$D_{k(\mathrm{Re}levance)} = \left[\frac{\sum_{i=1}^{n} w_i \cdot ABS\left(\frac{f_i^I - f_{i,k}^R}{f_{i,\max}^R - f_{i,\min}^R} \right)^2}{\sum_{i=1}^{n} w_i} \right]^{\frac{1}{2}} = \left[\frac{0.5 \cdot ABS\left(\frac{2-1}{4-1} \right)^2 + 0.5 \cdot ABS\left(\frac{1-2}{2-1} \right)^2}{0.5+0.5} \right]^{\frac{1}{2}}$$

$$D_{k(\mathrm{Re}levance)} = 0.745$$

As mentioned previously, when both the quality and relevance scores are considered when determining a case distance, they are weighted equally. For Case Id 296, case distance can be evaluated using Equation 3.2 as below:

$$D_k = \frac{1}{2} \cdot \left(D_{k(Quality)} + D_{k(\mathrm{Re}levance)} \right) = \frac{1}{2}(0.667 + 0.745) = 0.706$$

Using the same approach, case distances are computed for all of the retrieved cases and the results are as listed in Table 8.3. Furthermore, Table 8.3 also presents the results (CRFs) extracted from the retrieved cases, however, in this example, only CRFs for total collisions are considered.

Table 8.1. Example: Retrieved Cases and their Quality Scores

Retrieved Case	Study	Treatment of the Confounding Factors				
(Case Id)	(Study Id)	Changes in Traffic Volume	Inclusion of Unrelated Effects	RTM Artifact	Total # of Factors Treated	Quality Score, $D_{k(Quality)}$
169	54	No	No	No	0	1.000
192	59	Yes	No	No	1	0.667
238	68	Yes	No	No	1	0.667
257	73	No	No	No	0	1.000
267	74	No	No	No	0	1.000
296	77	No	Yes	No	1	0.667
297	77	No	Yes	No	1	0.667
314	423	Yes	Yes	Yes	3	0.000
319	424	Yes	Yes	Yes	3	0.000
328	434	No	No	No	0	1.000
359	448	No	No	No	0	1.000
406	101	No	No	No	0	1.000
407	101	No	No	No	0	1.000
408	101	No	No	No	0	1.000
463	116	No	No	No	0	1.000
480	119	No	Yes	No	1	0.667
498	124	No	No	No	0	1.000
689	78	Yes	No	No	1	0.667
690	78	Yes	No	No	1	0.667
713	196	No	No	No	0	1.000
766	1	No	No	No	0	1.000
796	229	No	No	No	0	1.000
835	376	No	No	No	0	1.000
854	403	No	Yes	No	1	0.667
855	403	No	Yes	No	1	0.667
856	403	No	Yes	No	1	0.667
861	254	No	No	No	0	1.000
869	256	Yes	Yes	Yes	3	0.000
870	257	No	Yes	No	1	0.667
885	262	No	Yes	No	1	0.667
893	265	No	No	No	0	1.000
894	265	No	No	No	0	1.000
895	265	No	No	No	0	1.000
896	265	No	No	No	0	1.000
897	265	No	No	No	0	1.000
898	265	No	No	No	0	1.000
899	265	No	No	No	0	1.000

Table 8.1. Example: Retrieved Cases and their Quality Scores (cont.)

Retrieved Case	Study	Treatment of the Confounding Factors				
(Case Id)	(Study Id)	Changes in Traffic Volume	Inclusion of Unrelated Effects	RTM Artifact	Total # of Factors Treated	Quality Score, $D_{k(Quality)}$
901	267	No	No	No	0	1.000
902	267	No	No	No	0	1.000
903	267	No	No	No	0	1.000
923	27	No	No	No	0	1.000
925	89	No	No	No	0	1.000
930	282	No	No	No	0	1.000
942	295	No	No	No	0	1.000
944	297	No	No	No	0	1.000
958	311	No	No	No	0	1.000
1253	292	No	No	No	0	1.000
2800	1	No	No	No	0	1.000
2805	285	No	No	No	0	1.000
2820	409	Yes	Yes	Yes	3	0.000
2828	413	No	No	No	0	1.000
2867	493	No	No	No	0	1.000
2872	496	No	No	No	0	1.000

Table 8.2. Example: Retrieved Cases and their Relevance Scores

Retrieved Case	Study	Location Characteristics			
(Case Id)	(Study Id)	Area Type	Implementation Level	Traffic Volume (AADT)	Relevance Score, $D_{k(Relevance)}$
169	54	Other	Wide Area		0.850
192	59	Other	Wide Area		0.850
238	68	Other	Wide Area		0.850
257	73	Other	Wide Area		0.850
267	74	Other	Wide Area		0.850
296	77	Urban	Wide Area		0.745
297	77	Urban	Wide Area		0.745
314	423	Rural	Wide Area		0.745
319	424	Other			0.667
328	434	Urban	Wide Area		0.745
359	448	Other	Wide Area		0.850
406	101	Rural	Wide Area		0.745
407	101	Rural	Wide Area		0.745
408	101	Urban	Wide Area		0.745
463	116	Urban			0.333
480	119	Urban			0.333
498	124	Urban			0.333
689	78	Urban	Wide Area		0.745
690	78	Urban	Wide Area		0.745
713	196	Urban			0.333
766	1	Rural	Wide Area		0.745
796	229	Rural	Wide Area		0.745
835	376	Other	Wide Area		0.850
854	403	Urban	Wide Area		0.745
855	403	Rural	Wide Area		0.745
856	403	Other	Wide Area		0.850
861	254	Urban	Wide Area		0.745
869	256	Rural	Wide Area		0.745
870	257	Rural	Wide Area		0.745
885	262	Urban	Wide Area		0.745
893	265	Other	Wide Area		0.850
894	265	Other	Wide Area		0.850
895	265	Other	Wide Area		0.850
896	265	Other	Wide Area		0.850
897	265	Other	Wide Area		0.850
898	265	Urban	Wide Area		0.745
899	265	Rural	Wide Area		0.745
901	267	Other	Wide Area		0.850

Table 8.2. Example: Retrieved Cases and their Relevance Scores (cont.)

Retrieved Case	Study	Location Characteristics			
(Case Id)	(Study Id)	Area Type	Implementation Level	Traffic Volume (AADT)	Relevance Score, $D_{k(Relevance)}$
902	267	Other	Wide Area	5000-9999	0.715
903	267	Other	Wide Area		0.850
923	27	Other	Isolated Location	20000-29999	0.385
925	89	Urban	Isolated Location		0.236
930	282	Other	Wide Area		0.850
942	295	Other			0.667
944	297	Other	Wide Area		0.850
958	311	Other	Wide Area		0.850
1253	292	Other	Wide Area		0.850
2800	1	Rural	Wide Area		0.745
2805	285	Other	Wide Area		0.850
2820	409	Other	Wide Area		0.850
2828	413	Other	Wide Area		0.850
2867	493	Other	Wide Area		0.850
2872	496	Other			0.667

Table 8.3. Example: Retrieved Cases (Ranked) and their Case Distances and Results

Case Id	Study Id	Quality Score, $D_{k(Quality)}$	Relevance Score, $D_{k(Relevance)}$	Case Distance, D_k	$(1-D_k)$	Case Result, CRF_k (%)	Weighted Case Result (%)
169	54	1.000	0.850	0.925	0.075	15.00	1.13
192	59	0.667	0.850	0.758	0.242	25.00	6.04
238	68	0.667	0.850	0.758	0.242	79.00	19.10
257	73	1.000	0.850	0.925	0.075	40.00	3.00
267	74	1.000	0.850	0.925	0.075	60.00	4.50
296	77	0.667	0.745	0.706	0.294	75.13	22.09
297	77	0.667	0.745	0.706	0.294	45.28	13.31
314	423	0.000	0.745	0.373	0.627	22.00	13.80
319	424	0.000	0.667	0.333	0.667	5.00	3.33
328	434	1.000	0.745	0.873	0.127	57.00	7.26
359	448	1.000	0.850	0.925	0.075	30.00	2.25
406	101	1.000	0.745	0.873			
407	101	1.000	0.745	0.873	0.127	30.00	3.82
408	101	1.000	0.745	0.873			
463	116	1.000	0.333	0.667			
480	119	0.667	0.333	0.500	0.500	9.00	4.50
498	124	1.000	0.333	0.667	0.333	30.00	10.00
689	78	0.667	0.745	0.706	0.294	58.00	17.05
690	78	0.667	0.745	0.706			
713	196	1.000	0.333	0.667	0.333	30.00	10.00
766	1	1.000	0.745	0.873	0.127	58.00	7.38
796	229	1.000	0.745	0.873	0.127	30.00	3.82
835	376	1.000	0.850	0.925	0.075	21.00	1.58
854	403	0.667	0.745	0.706	0.294	63.57	18.69
855	403	0.667	0.745	0.706	0.294	63.33	18.62
856	403	0.667	0.850	0.758	0.242	73.09	17.67
861	254	1.000	0.745	0.873			
869	256	0.000	0.745	0.373	0.627	10.00	6.27
870	257	0.667	0.745	0.706	0.294	20.00	5.88
885	262	0.667	0.745	0.706	0.294	50.00	14.70
893	265	1.000	0.850	0.925			
894	265	1.000	0.850	0.925			
895	265	1.000	0.850	0.925	0.075	30.00	2.25
896	265	1.000	0.850	0.925	0.075	21.00	1.58
897	265	1.000	0.850	0.925			
898	265	1.000	0.745	0.873	0.127	38.00	4.84
899	265	1.000	0.745	0.873	0.127	36.90	4.70
901	267	1.000	0.850	0.925	0.075	59.00	4.43
902	267	1.000	0.715	0.858	0.142	14.00	1.99
903	267	1.000	0.850	0.925	0.075	26.00	1.95

Table 8.3. Example: Retrieved Cases (Ranked) and their Case Distances and Results (cont.)

Case Id	Study Id	Quality Score, $D_{k(Quality)}$	Relevance Score, $D_{k(Relevance)}$	Case Distance, D_k	(1- D_k)	Case Result, CRF_k (%)	Weighted Case Result (%)
923	27	1.000	0.385	0.692			
925	89	1.000	0.236	0.618	0.382	-19.00	-7.26
930	282	1.000	0.850	0.925	0.075	24.69	1.85
942	295	1.000	0.667	0.833			
944	297	1.000	0.850	0.925			
958	311	1.000	0.850	0.925			
1253	292	1.000	0.850	0.925	0.075	-18.00	-1.35
2800	1	1.000	0.745	0.873	0.127	17.00	2.16
2805	285	1.000	0.850	0.925	0.075	69.00	5.18
2820	409	0.000	0.850	0.425			
2828	413	1.000	0.850	0.925			
2867	493	1.000	0.850	0.925	0.075	30.00	2.25
2872	496	1.000	0.667	0.833			
				Sum	**8.263**	**1327.99**	**260.38**

Table 8.3 illustrates again that cases with lower case distances are weighted more significantly than those with higher case distances. The non-weighted and weighted CRFs for total collisions can be determined by using Equations 3.3 and 3.4 respectively, as shown below:

$$CRF_{Non-Weighted} = \frac{\sum_{k=1}^{N} CRF_k}{N} = \frac{1327.99}{38} = 34.94\%$$

$$CRF_{Weighted} = \frac{\sum_{k=1}^{N} (1 - D_k) \cdot CRF_k}{\sum_{k=1}^{N} (1 - D_k)} = \frac{260.38}{8.263} = 31.51\%$$

Therefore, for the proposed countermeasure, i.e., general lighting improvement, the non-weighted and weighted CRFs are 34.94% and 31.51% respectively. As expected, the weighted CRF is less than the non-weighted CRF as with the weighted result, the confounding factors have been accounted for. Finally, the accuracy of these estimates can be represented by their standard deviations. The standard deviations for both the non-weighted and weighted CRFs can be determined with Equations 3.5 and 3.6 respectively, as shown below:

$$\sigma_{CRF-Non-Weighted} = \sqrt{\frac{\sum_{k=1}^{N} (CRF_k - CRF)^2}{N-1}} = \sqrt{\frac{21222.55}{38-1}} = 23.94\%$$

$$\sigma_{CRF-Weighted} = \sqrt{\frac{\sum_{k=1}^{N} (1 - D_k) \cdot (CRF_k - CRF)^2}{\sum_{k=1}^{N} (1 - D_k) \cdot (N-1)}} = \sqrt{\frac{5287.65}{305.73}} = 4.16\%$$

The standard deviations for the non-weighted and weighted CRFs are 23.94% and 4.16% respectively. Clearly, the non-weighted standard deviation is significantly greater than the weighted standard deviation.

8.3 Expected Collision Reduction

From the analysis conducted in Section 8.2, the expected values for the non-weighted and weighted CRFs are 34.94% and 31.51%, while the standard deviations are 23.94% and 4.16% respectively. Assuming that from the collision analysis conducted by the safety engineer, the expected number of collisions, N, for this location is 20.0 collisions per year (col/yr) with a standard deviation of 2.5 col/yr. Equations 6.1 and 6.2 can now be used to calculate the expected value and variance of Z respectively. For example, with the non-weighted results, the expected value of Z and its variance can be computed as follows:

$$E(Z) = N \times CRF = 20.0 \times 0.3494 = 6.99 col/yr$$

$$Var(Z) = N^2(\sigma_{CRF})^2 + CRF^2(\sigma_N)^2 + (\sigma_{CRF} \times \sigma_N)^2$$

$$= 20^2(0.2394)^2 + 0.3494^2(2.5)^2 + (0.2394 \times 2.5)^2 = 24.06(col/yr)^2$$

As for the weighted results, $E(Z) = 6.30 col/yr$ and $Var(Z) = 1.32(col/yr)^2$ can be determined in a similar fashion as shown above. The above results represent the expected reductions in total collisions and the uncertainties at the location.

8.4 Economic Analysis

With the results determined in Section 8.3, the expected value of BCR and its accuracy can now be evaluated. For the purpose of this example, the variable k in Equation 7.1 is assumed to be 0.25. Hence, the non-weighted BCR and its variance can be determined with Equations 7.1 and 7.3 respectively, as shown below:

$$E(BCR) = k \times E(Z) = 0.25 \times 6.99 = 1.747$$

$$Var(BCR) = k^2 \times Var(Z) = 0.25^2 \times 24.06 = 1.504$$

Similarly, the weighted BCR and its variance can be calculated and are determined to be 1.576 and 0.083 respectively. Therefore, given the expected value and variance of BCR, the two gamma distribution parameters, α and β corresponding to the non-weighted BCR can be determined using Equations 7.4 and 7.5 respectively as follows:

$$\beta = \frac{E(BCR)}{Var(BCR)} = \frac{1.747}{1.504} = 1.161$$

$$\alpha = E(BCR) \times \beta = 1.747 \times 1.161 = 2.03$$

Similarly, the α and β parameters for the weighted BCR are 30.017 and 19.051 respectively. It follows that with the known gamma parameters, a cumulative distribution can be plotted as shown in Figure 8.1 (see Appendix E for the probability density plot). Subsequently, the probability of achieving a specific economic goal can be evaluated prior to the implementation of a countermeasure. For example, if it is desired to determine the probability of achieving a BCR of 2.0 with the proposed countermeasure (general lighting improvement), Figure 8.1 indicates that the probabilities of achieving this goal (or having a BCR of less or equal to 2.0) are 61.5% and 80.6% based on the non-weighted and weighted BCR curves respectively. Based on Figure 8.1, it is clear that the expected range of BCR represented by the weighted BCR curve is considerably smaller than the one indicated by the non-weighted curve, i.e., a BCR range of 1.0 to 2.5 compared to a range of 0.0 to 6.0 respectively.

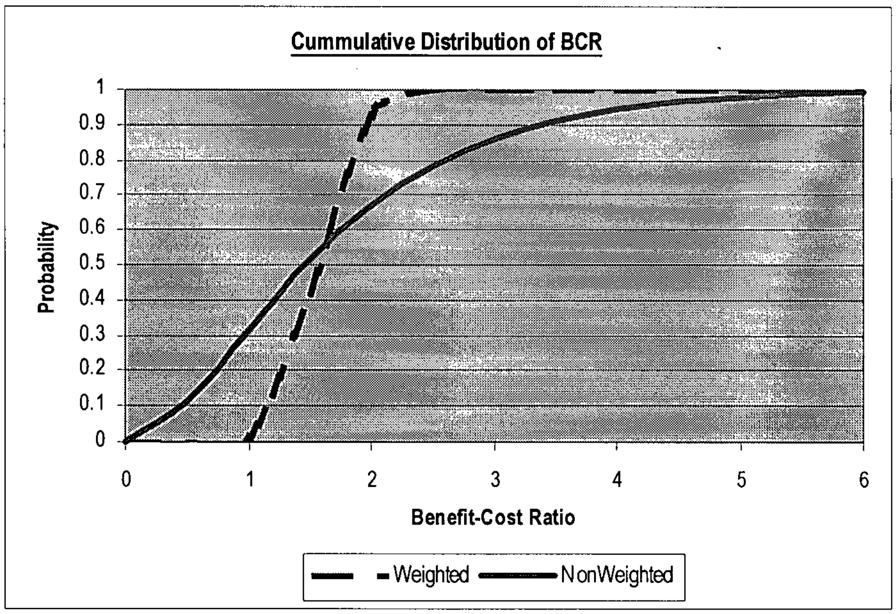


Figure 8.1. Example: Cumulative Distribution Plot of BCR

9.0 CONCLUSION

This thesis first describes the development of ISECR, which is now a functional intelligent database on the MS Access platform. ISECR maintains a case base that consists of published literature which quantifies crash reduction benefits for various safety improvements. Utilizing a CBR approach, ISECR permits users to query the system for cases similar to the current situation, retrieves and then summarizes the retrieved solutions to estimate the range and reliability of the countermeasure effectiveness on a project level.

With the ISECR outputs, this thesis then illustrates the use of the moment approach to determine the expected collision reduction and its uncertainties for specific countermeasures. Lastly, the technique involving the probability assessment of achieving a specific benefit-cost ratio for a specific countermeasure is also presented in the thesis. The evaluation procedure illustrated in this report attempts to increase the confidence of predictions when evaluating the expected benefits from safety improvement programs.

Currently, 450 evaluation studies are entered into the ISECR case base. To further increase the usefulness and applicability of ISECR, new evaluation studies should be evaluated and entered into the database on a regular basis. Based on the available cases in the database, ISECR has shown to provide results that are comparable to the results obtained from real cases, given that there is an adequate sample size of retrieved studies that matches the user's query. Nevertheless, further data testing and validation of the system from experts and end-users are still necessary to improve the prototype ISECR.

The confounding factors considered in ISECR can be further expanded to include other factors, such as collision migration. A different weighting scheme for the confounding factors can also be introduced to highlight the relative importance of each confounding factor. The same can be applied to the location characteristics considered in this thesis, i.e., other location characteristics, in addition to the ones included in ISECR, can also be introduced and a different weighting scheme can also be employed to emphasize the importance of some characteristics.

In addition, the prototype ISECR can be improved by providing its user the capability of predicting CRFs for different combinations of countermeasures, instead of one countermeasure as currently allowed in ISECR. Frequently, more than one countermeasure is considered at a location, i.e., intersection improvements may include a new left-turn lane, new delineation, lane widening. Thus, it would definitely be beneficial to be able to query for the effect of a combination of countermeasures at a location.

Additional risk analysis should also be performed for the economic procedure illustrated in this thesis. This is essential as the discount rate, *i*, and project life, *t*, used in Equation 7.1 were assumed to be constant over time. Nevertheless, these two variables can fluctuate during the life of the project and are subject to changes in interest rate, risk premium, market condition, government policy, etc. Hence, further risk analysis incorporating the random nature of these two variables should be conducted.

The computerized approach which ISECR employs minimizes the amount of manual work required for safety analysts to determine the effectiveness of safety improvements. However, it should be noted that ISECR is not intended to eliminate the use of engineering. In many cases, the safety analyst has to use his/her judgement in evaluating the results produced by ISECR.

BIBLIOGRAPHY

Aamodt, A. and Plaza, E. (1994). *Case-based reasoning: foundational issues, methodological variations, and system approaches*, Artificial Intelligence Communications, Vol.7, No.1, pp.39-59.

Ang, A. and Tang, W. (1984). *Probability concepts in engineering planning and design*, John Wiley and Sons, Toronto, Canada.

Benjamin, J. and Cornell, C. (1970). *Probability, statistics, and decision for civil engineers*, McGraw-Hill, Toronto, Canada.

Brude, U. and Larsson, J. (1988). *The use of prediction models for eliminating effects due to regression-to-the-mean in road accident data*, Accident Analysis and Prevention, Vol.20, No.4, pp.299-310.

Capus, L. and Tourigny, N. (1998). *Road safety analysis: a case-based reasoning approach*, Transportation Research Board 77th Annual Meeting, Washington, D.C.

Creasey, T. and Agent, K.R. (1985). *Development of accident reduction factors*, Research Report UKTRP-85-6, Lexington, KY, Kentucky Transportation Cabinet, Federal Highway Administration.

Cribbins, P. and Walton, C. (1970). *Traffic signals and overhead flashers at rural intersections: their effectiveness in reducing accidents*, Highway Research Record, Vol.325, pp.1-14.

Elvik, R. (1995). *Meta-analysis of evaluations of public lighting as accident countermeas*ure, Transportation Research Record, No.1485, pp.112-123.

Elvik, R. (1996). *Evaluations of road accident blackspot treatment: a case of the iron law of evaluation studies?*, Accident Analysis and Prevention, Vol.28, No.6, pp.685-694.

Elvik, R. (1998). *Are road safety evaluation studies published in peer reviewed journals more valid than similar studies not published in peer reviewed journals?*, Accident Analysis and Prevention, Vol.30, No.1, pp.101-118.

Elvik, R. (1998). *Evaluating the statistical conclusion validity of weighted mean results in meta-analysis by analyzing funnel graph diagrams*, Accident Analysis and Prevention, Vol.30, No.2, pp.255-266.

Gonzalez, A.J. and Laureano-Ortiz, R. (1992). *A case-based reasoning approach to real estate property appraisal, Expert Systems with Applications*, Vol.4, pp.229-246.

Hauer, E. (1992). *Empirical Bayes approach to the estimation of "unsafety": the multivariate regression method*, Accident Analysis and Prevention, Vol.24, No.5, pp.457-477.

Hauer, E., Ng, J., and Lovell, J. (1988). *Estimation of safety at signalized intersections*, Transportation Research Record, No.1185, pp.48-61.

Jovanis, P. and Chang, H. (1986). *Modeling the relationship of accidents to miles traveled*, Transportation Research Record, No.1068, pp.42-51.

Khattak, A. and Kanafani, A. (1996). *Case-based reasoning: a planning tool for intelligent transportation systems*, Transportation Research Part C-Emerging Technologies, Vol.4, No.5, pp.267-288.

Khattak, A. and Renski, H. (1999). *Plan HOV: a case-based reasoning planning tool for high-occupancy-vehicle lane analysis in a GIS environment*, Transportation Research Board 78th Annual Meeting, Washington D.C.

Kolodner, J. (1993). *Case-based reasoning*, Morgan Kaufmann Publishers, Inc.

Kulmala, R. (1995). *Safety at rural three- and four-arm junctions. Development of accident prediction models*, Espoo 1995, Technical Research Centre of Finland, VTT 233.

Leake, D.B. and Plaza E. (1997). *Case-based reasoning research and development: second international conference on case-based reasoning*, Providence, RI, USA.

Maher, M.L., Balachandran, M.B., and Zhang, D.M. (1995). *Case-based reasoning in design*, Lawrence Erlbaum Associates.

Mayer, P. (1971). *Relating traffic control and roadway elements to highway safety- the relationship between highway transportation and safety*, Traffic Engineering, pp.23-27.

McFarland, W.F., Griffin, L.I., Rollins, J.B., Stockton, W.R., Phillips, D.T., and Dudek, C.L. (1978). *Assessment of techniques for cost-effective of highway accident countermeasures*, Report No. FHWA-RD-79-53, Washington, D.C., Federal Highway Administration.

Miaou, S. and Lum, H. (1993). *Modeling vehicle accident and highway geometric design relationships*, Accident Analysis and Prevention, Vol.25, No.6, pp.689-709.

Pu, P. (1993). *Introduction: issues in case-based design systems*, AI EDAM, Vol.7, No.2, pp.79-85.

Pu, P. and Maher, M. (1998). *Issues and applications of case-based reasoning in design*, Lawrence Erlbaum Associates.

Saccomanno, F. and Buyco, C. (1988). *Generalized log-linear models of truck accident rates*, Paper presented at Transportation Research Board 67th Annual Meeting, Washington, D.C.

Schank, R.C. (1982). *Dynamic memory: a theory of reminding and learning in computers and people*, Cambridge University Press, Cambridge, England.

Tamburri, T.N. and Smith, R.N. (1971). *The safety index: a method of evaluating and rating safety benefits*, Highway Research Record 332, pp.28-43.

Terry, D.A. and Watson, J.E. (1982). *Post-implementation evaluation system methodology and output*, Traffic and Safety Division, New York State Department of Transportation.

Watson, I. (1997). *Applying case-based reasoning: techniques for enterprise systems*, Morgan Kaufmann Publishers, Inc.

Wilson, J. (1967). *Simple types of intersection improvements*, Highway Research Board, Special Report 93, pp.144-159.

Yeh, I.C. (1997). *Case-based approaches for preliminary design of steel building frames*, Microcomputers in Civil Engineering, Vol.12, pp.327-337.

APPENDIX A: COUNTERMEASURE TYPES

Table A.1. Countermeasure Types

Countermeasure Category	Countermeasure Type
1. Area-wide schemes	General
	Blackspot treatment
	Enforcement
	New roads
	Traffic calming: general
	Traffic calming: speed bumps/humps
	Traffic planning/Management
2. Bridge improvements	General
	Widen bridges
3. Cyclist/Pedestrian facilities	General bicycle safety improvements
	General pedestrian safety improvements
	Pedestrian crossings
	Pedestrian overpasses/underpasses
	Pedestrian signals
4. Delineation	General
	Pavement markings: general
	Pavement markings: edge lines
	Pavement markings: median edge line
	Pavement markings: right edge line
	Raised pavement markers
	Reflected guide posts
	Strips: general
	Strips: rumble
	Strips: transverse
5. Geometric improvements	General
	Alignment: general
	Alignment: horizontal
	Alignment: vertical
	Median: general
	Median: close median openings
	Median: concrete/attenuator barriers
	Median: install new median
	Median: upgrade existing median
	Median: widen existing median
	Sight distance
	Staggered intersection
	Superelevation

Table A.1. Countermeasure Types (cont.)

Countermeasure Category	Countermeasure Type
6. Intersection improvements	General
	Channelization: general
	Channelization: left-turn lane: general
	Channelization: left-turn lane with left-turn phase
	Channelization: left-turn lane without left-turn phase
	Channelization: right-turn lane
	Turning bay/traffic island
	Roundabouts: general
	Roundabouts: installation/upgrade
7. Lane/Shoulder treatment	General
	2-way left-turn lane
	Acc./Decel./Passing lane
	Bicycle lane
	Bus lane
	Climbing lane
	Flattening side-slopes
	HOV lane
	Lane/shoulder: general
	Lane/shoulder: narrowing
	Lane/shoulder: widening
	Lane addition
8. Lighting improvements	General
	Install new lighting
	Upgrade existing lighting
9. Object removal/relocation	General
	Relocation: general
	Relocation: fixed objects
	Relocation: utility poles
	Removal: general
	Removal: fixed objects
	Removal: trees
	Removal: utility poles
10. Parking improvements	General
	Change from angle- to parallel parking
	Eliminate parking: general
	Eliminate parking: parallel parking
11. Pavement treatment	General
	Pavement grooving
	Resurfacing
	Skid reduction

Table A.1. Countermeasure Types (cont.)

Countermeasure Category	Countermeasure Type
12. Railway improvements	General
	Automatic gates
	Flashing beacons
13. Regulation change	General
	Modify speed limit: general
	Modify speed limit: decrease
	Modify speed limit: increase
	Prohibit turns: general
	Prohibit turns: left-turns
14. Safety barriers	General
	Crash cushions
	Guardrails: general
	Guardrails: double-sided
	Safety poles/posts
15. Traffic controls/signs	General
	Guidance signs
	Install new signs/upgrade existing signs
	Regulatory signs: general
	Regulatory signs: speed
	Regulatory signs: stop signs: general
	Regulatory signs: stop signs: 2-way to 4-way stops
	Regulatory signs: stop signs: 4-way stops
	Regulatory signs: minor-leg stops
	Regulatory signs: yield signs
	Warning signs: general
	Warning signs: flashing beacons/signals
16. Traffic signals	General
	Actuated signals
	Advance warning signs
	Coordinated signals
	Flashing beacons/signals: general
	Flashing beacons/signals: all-way red
	Flashing beacons/signals: red-yellow
	Install new signals/upgrade existing signals
	Phasing: general
	Phasing: left-turn phase
	Phasing: pedestrian phase
	Phasing: timing
	Removal signals

APPENDIX B: LOCATION CHARACTERISTICS CONSIDERED FOR EACH OF THE EIGHT LOCATION TYPES

Table B.1. Representation of Location Characteristics for General Intersections

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Intersection type	1=four-legged 2=T-intersection 3=Y-intersection
Implementation level	1=isolated location 2=wide area
Total entering traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Provision of left-turn channelization	1=yes 2=no
Provision of right-turn channelization	1=yes 2=no
Left-turn movement	1=not allowed 2=permissive 3=protected
Right-turn movement	1=not allowed 2=permissive 3=protected
Street parking	1=allowed 2=not allowed
Average number of lanes per approach	1=less than or equal to 2 lanes 2=more than 2 lanes

Table B.2. Representation of Location Characteristics for Signalized Intersections

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Intersection type	1=four-legged 2=T-intersection 3=Y-intersection
Implementation level	1=isolated location 2=wide area
Total entering traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Provision of left-turn channelization	1=yes 2=no
Provision of right-turn channelization	1=yes 2=no
Left-turn movement	1=not allowed 2=permissive 3=protected
Right-turn movement	1=not allowed 2=permissive 3=protected
Street parking	1=allowed 2=not allowed
Type of traffic control	1=fixed-timed 2=semi-actuated 3=fully-actuated
Average number of lanes per approach	1=less than or equal to 2 lanes 2=more than 2 lanes

Table B.3. Representation of Location Characteristics for Unsignalized Intersections

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Intersection type	1=four-legged 2=T-intersection 3=Y-intersection
Implementation level	1=isolated location 2=wide area
Total entering traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Provision of left-turn channelization	1=yes 2=no
Provision of right-turn channelization	1=yes 2=no
Left-turn movement	1=not allowed 2=permissive 3=protected
Right-turn movement	1=not allowed 2=permissive 3=protected
Street parking	1=allowed 2=not allowed
Type of traffic control	1=uncontrolled 2=2-way stops 3=4-way stops
Average number of lanes per approach	1=less than or equal to 2 lanes 2=more than 2 lanes

Table B.4. Representation of Location Characteristics for Road Sections

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Implementation level	1=isolated location 2=wide area
Total traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Provision of left-turn channelization	1=yes 2=no
Provision of right-turn channelization	1=yes 2=no
Street parking	1=allowed 2=not allowed
Average number of lanes	1=less than or equal to 2 lanes 2=more than 2 lanes
Passing/Acceleration/ Deceleration Lanes	1=yes 2=no

Table B.5. Representation of Location Characteristics for Freeways

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Implementation Level	1=isolated location 2=wide area
Total traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Average number of lanes	1=less than or equal to 2 lanes 2=more than 2 lanes
Passing/Acceleration/ Deceleration Lanes	1=yes 2=no

Table B.6. Representation of Location Characteristics for Bridges, Rails, and Construction Zones

Characteristic	Feature Value
Area type	1=urban 2=suburban 3=rural 4=other
Implementation Level	1=isolated location 2=wide area
Total traffic volume (AADT)	1=0-4999 2=5000-9999 3=10000-14999 4=15000-19999 5=20000-29999 6=30000-39999 7=40000-49999 8=50000-59999 9=60000-69999 10=70000-79999 11=80000 and more
Average lane width	1=less than 12 ft 2=greater or equal to 12 ft
Average number of lanes	1=less than or equal to 2 lanes 2=more than 2 lanes

APPENDIX C: ISECR WINDOWS: COUNTERMEASURE TYPES CONSIDERED FOR DIFFERENT COUNTERMEASURE CATEGORIES

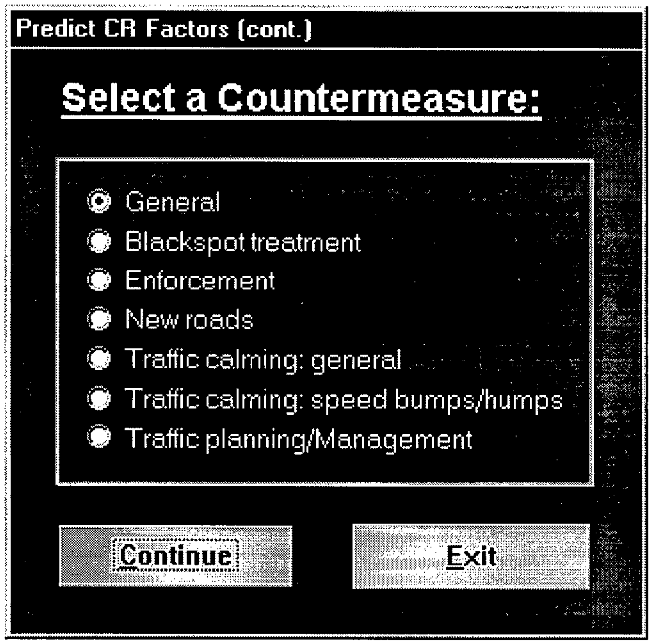


Figure C.1. ISECR Window: Countermeasure Types for Area-Wide Schemes

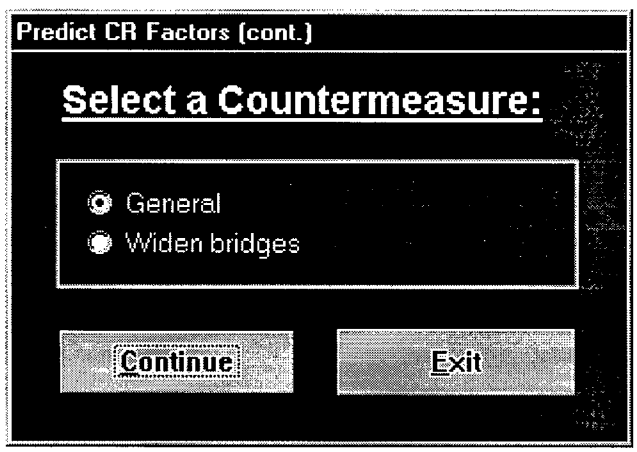


Figure C.2. ISECR Window: Countermeasure Types for Bridge Improvements

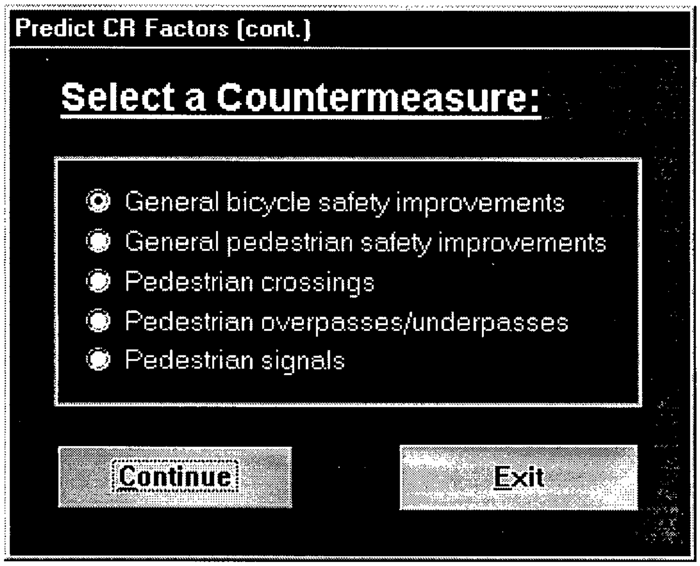


Figure C.3. ISECR Window: Countermeasure Types for Cyclist/Pedestrian Facilities

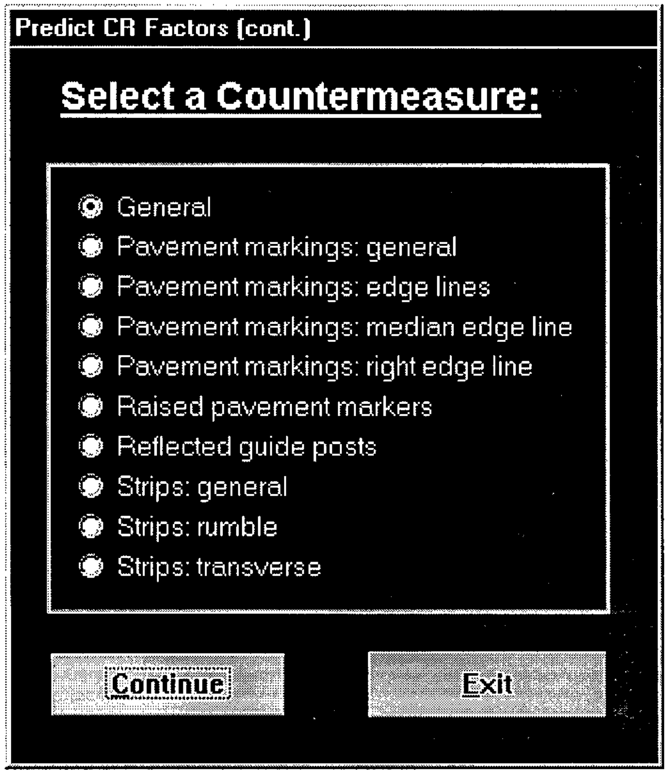


Figure C.4. ISECR Window: Countermeasure Types for Delineation

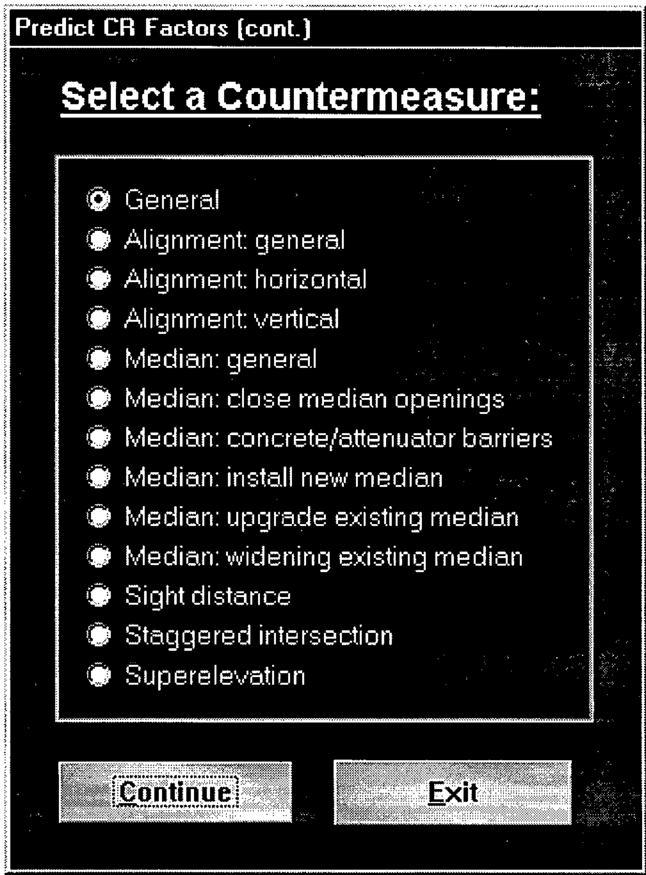


Figure C.5. ISECR Window: Countermeasure Types for Geometric Improvements

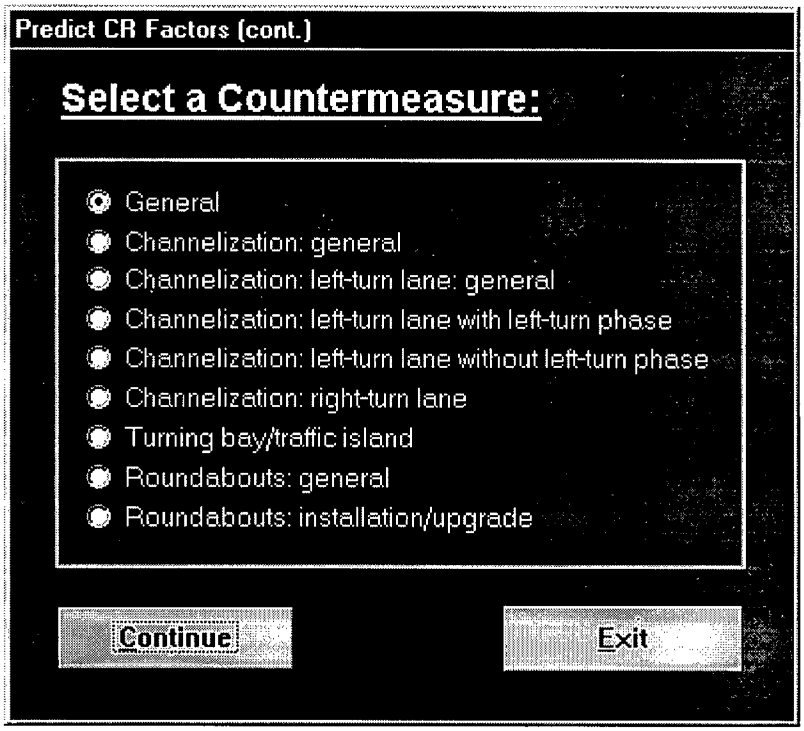


Figure C.6. ISECR Window: Countermeasure Types for Intersection Improvements

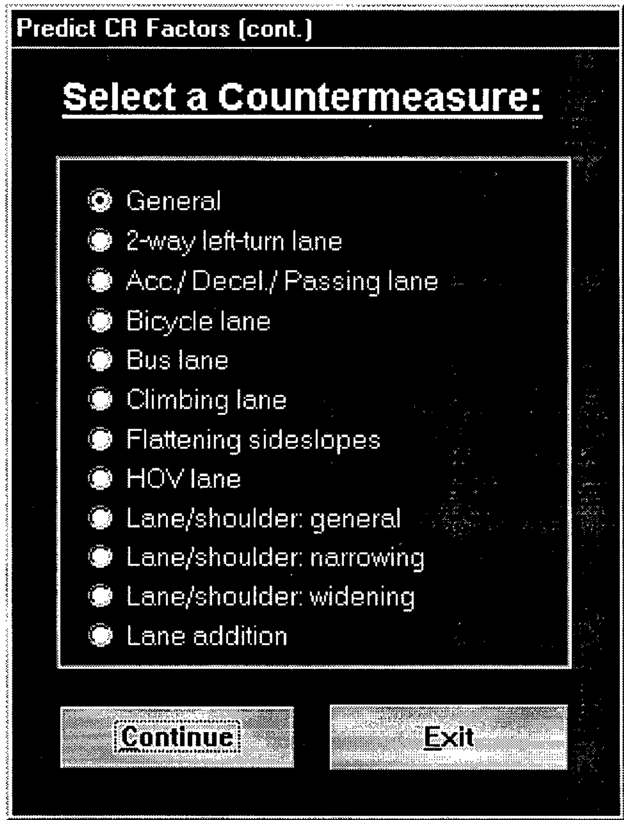


Figure C.7. ISECR Window: Countermeasure Types for Lane/Shoulder Treatment

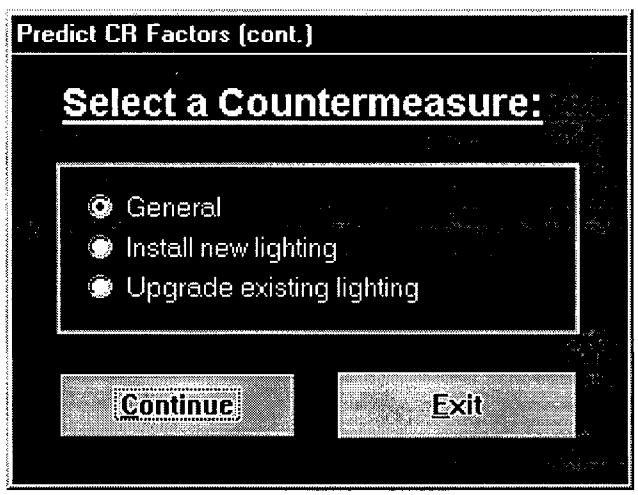


Figure C.8. ISECR Window: Countermeasure Types for Lighting Improvements

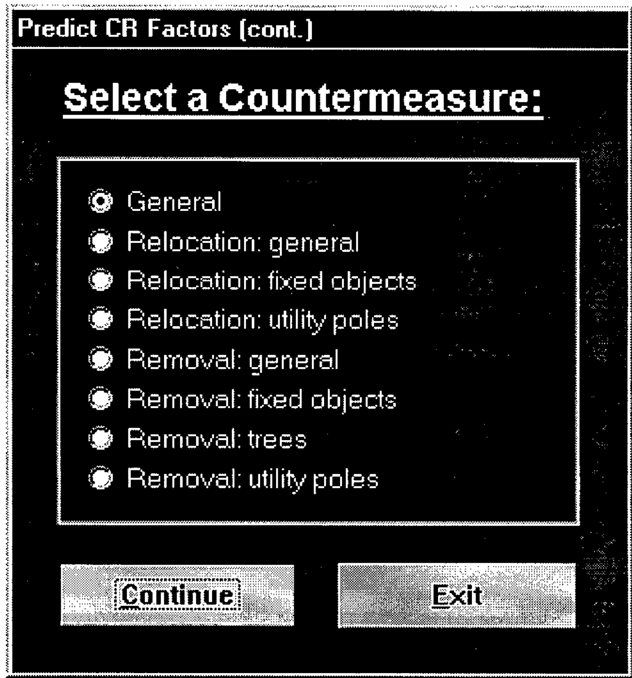


Figure C.9. ISECR Window: Countermeasure Types for Object Removal/Relocation

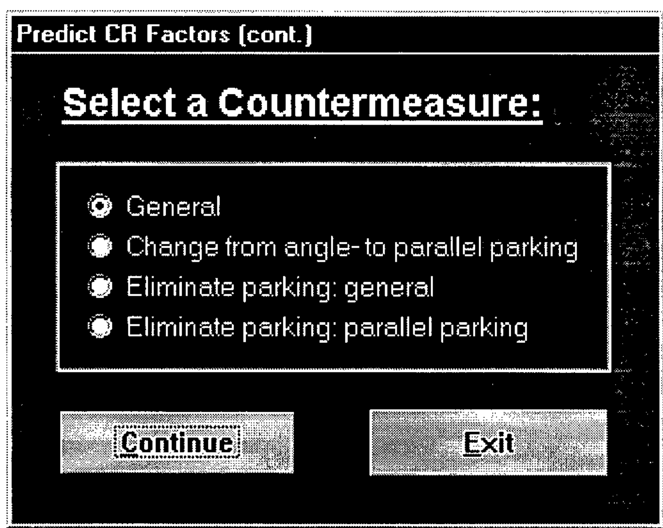


Figure C.10. ISECR Window: Countermeasure Types for Parking Improvements

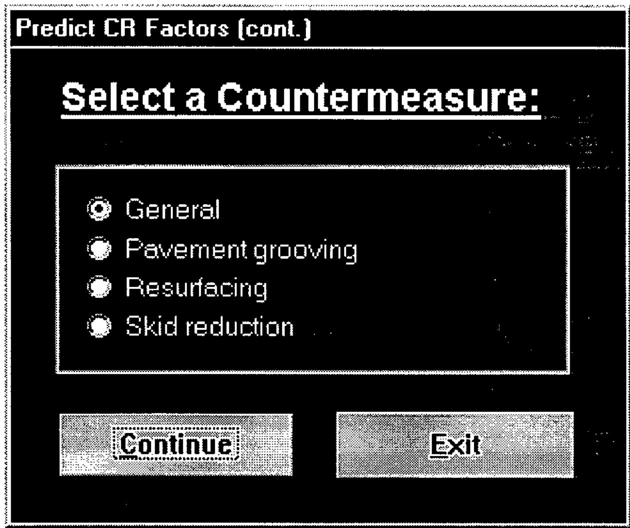


Figure C.11. ISECR Window: Countermeasure Types for Pavement Treatment

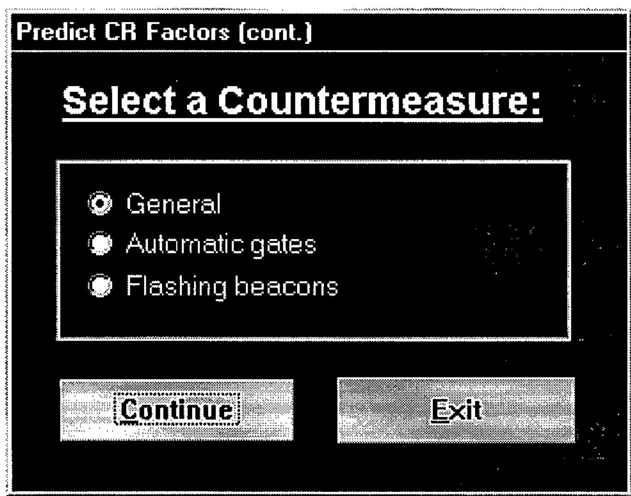


Figure C.12. ISECR Window: Countermeasure Types for Railway Improvements

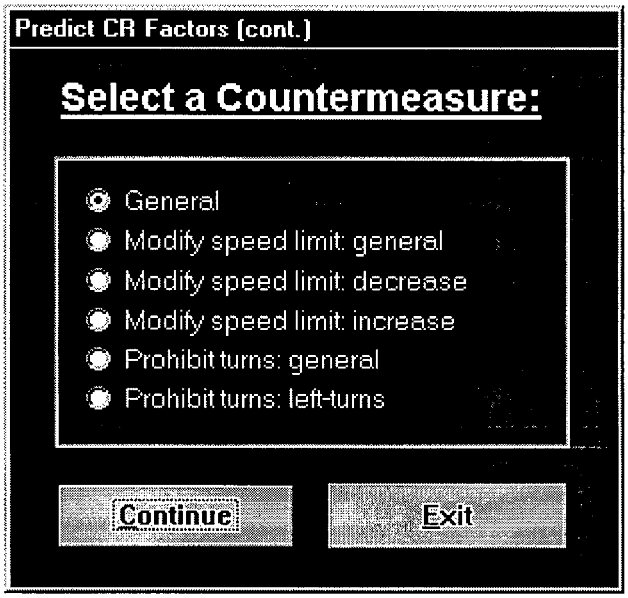


Figure C.13. ISECR Window: Countermeasure Types for Regulation Change

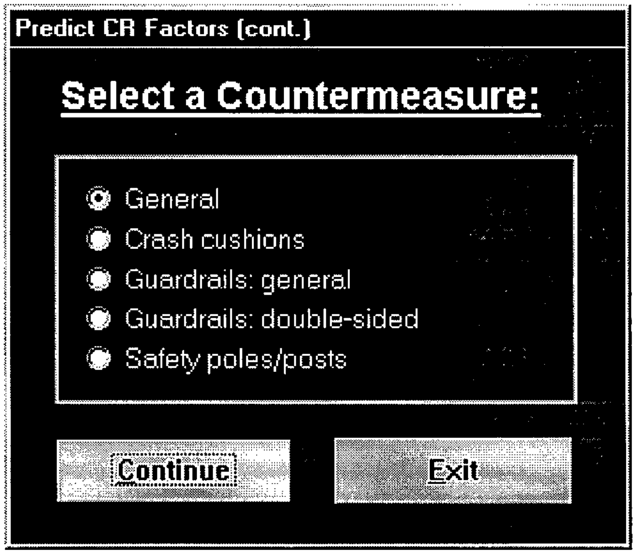


Figure C.14. ISECR Window: Countermeasure Types for Safety Barriers

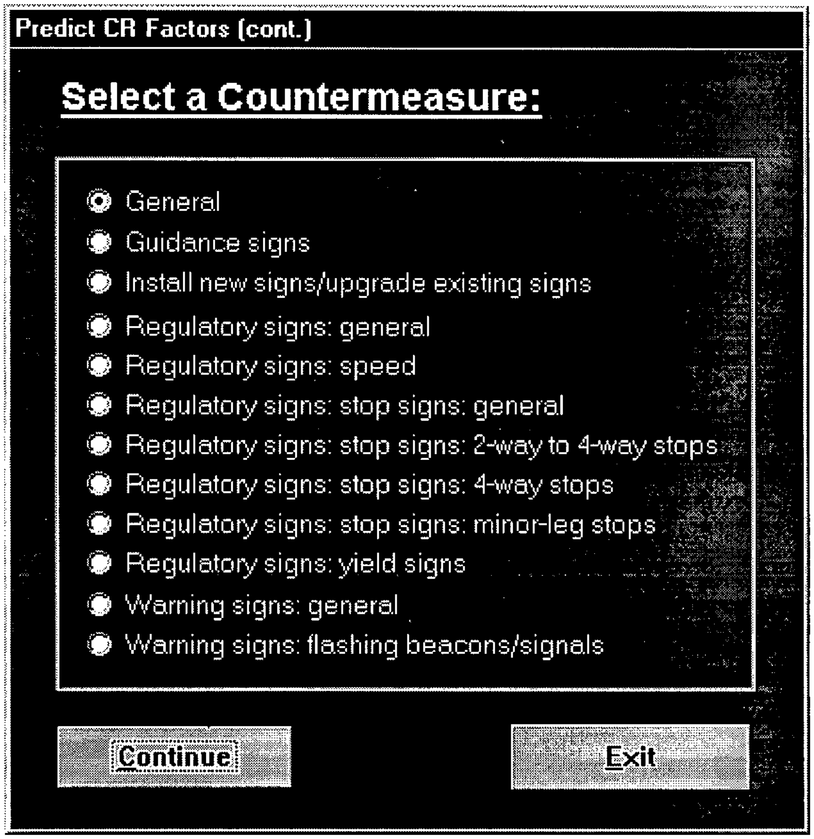


Figure C.15. ISECR Window: Countermeasure Types for Traffic Controls/Signs

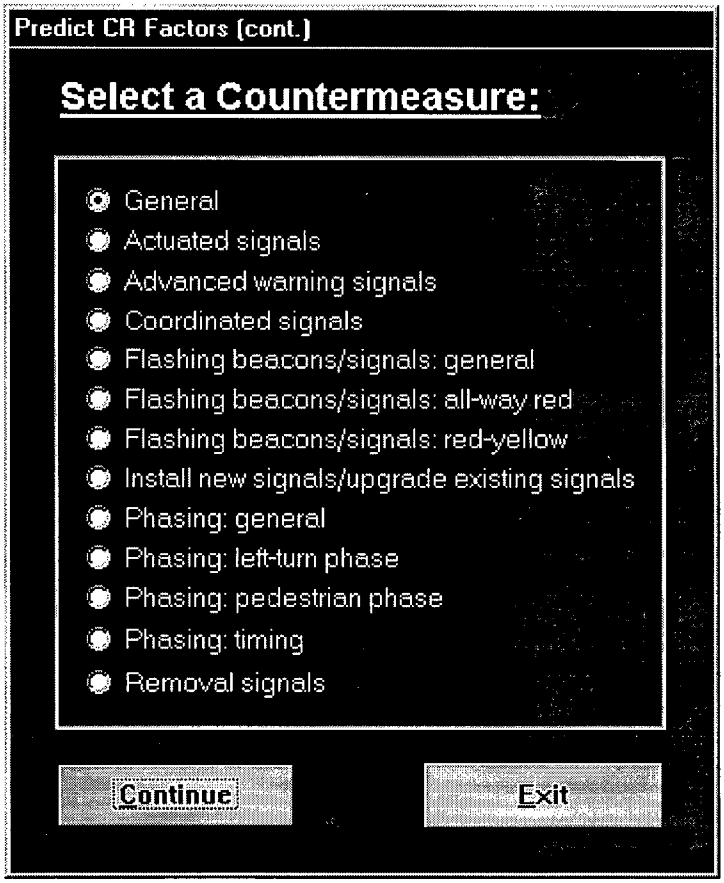


Figure C.16. ISECR Window: Countermeasure Types for Traffic Signals

APPENDIX D: ISECR WINDOWS: INPUT FORMS USED TO ENTER LOCATION CHARACTERISTICS FOR DIFFERENT LOCATION TYPES

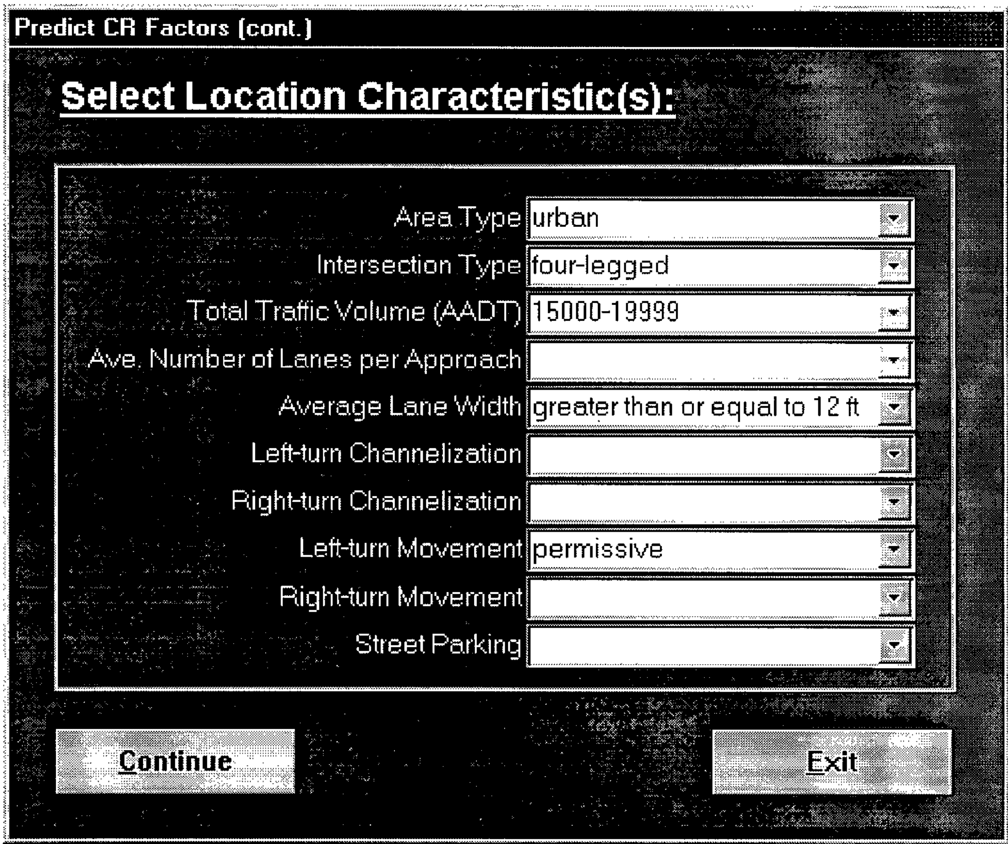


Figure D.1. ISECR Window: Input Location Characteristics for General Intersections

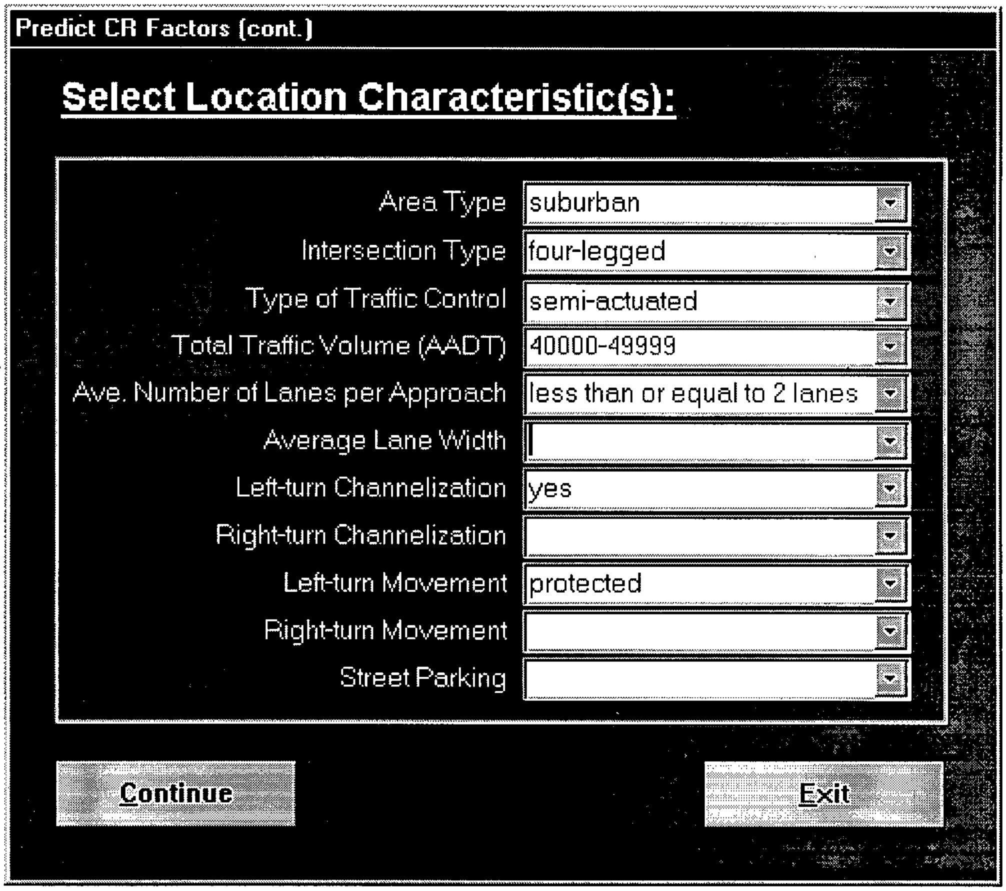


Figure D.2. ISECR Window: Input Location Characteristics for Signalized Intersections

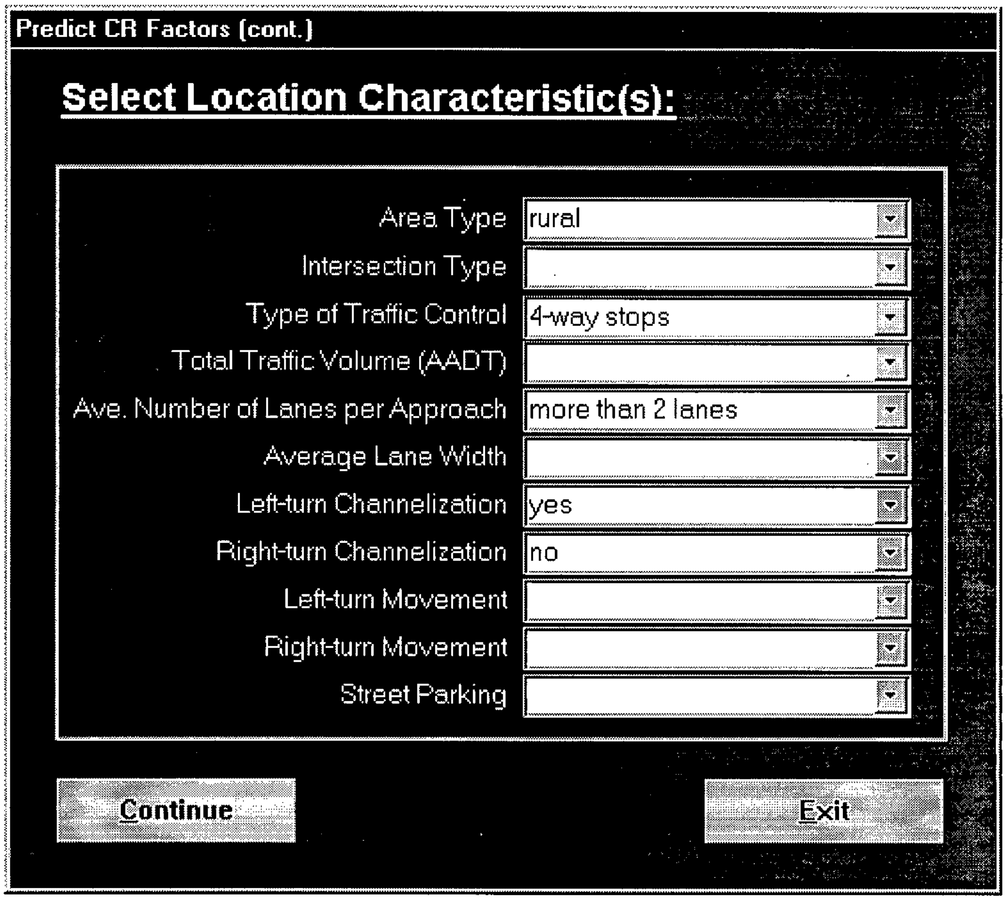


Figure D.3. ISECR Window: Input Location Characteristics for Unsignalized Intersections

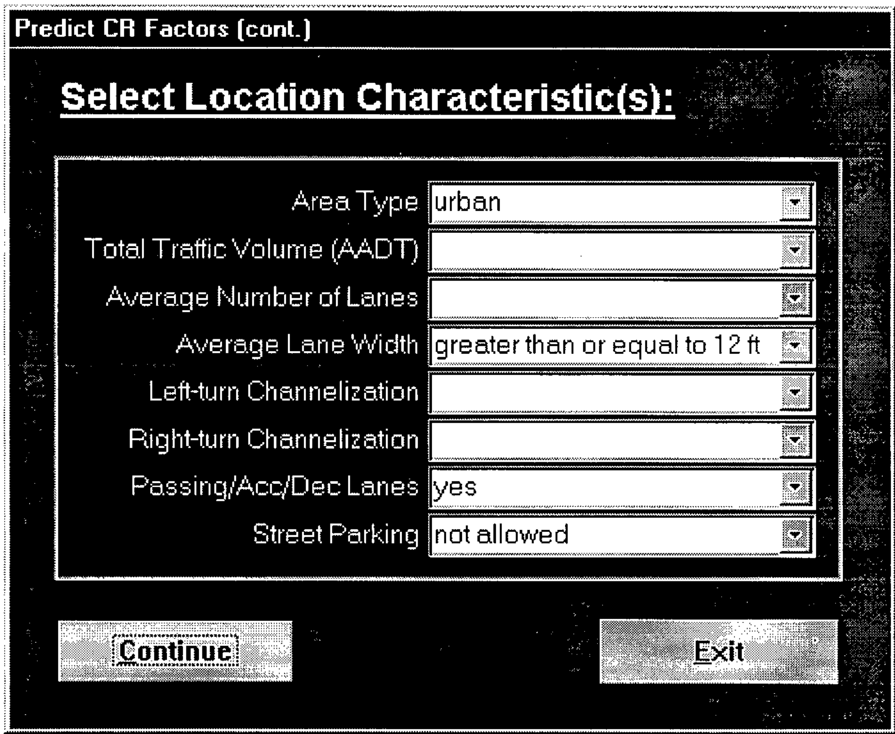


Figure D.4. ISECR Window: Input Location Characteristics for Road Sections

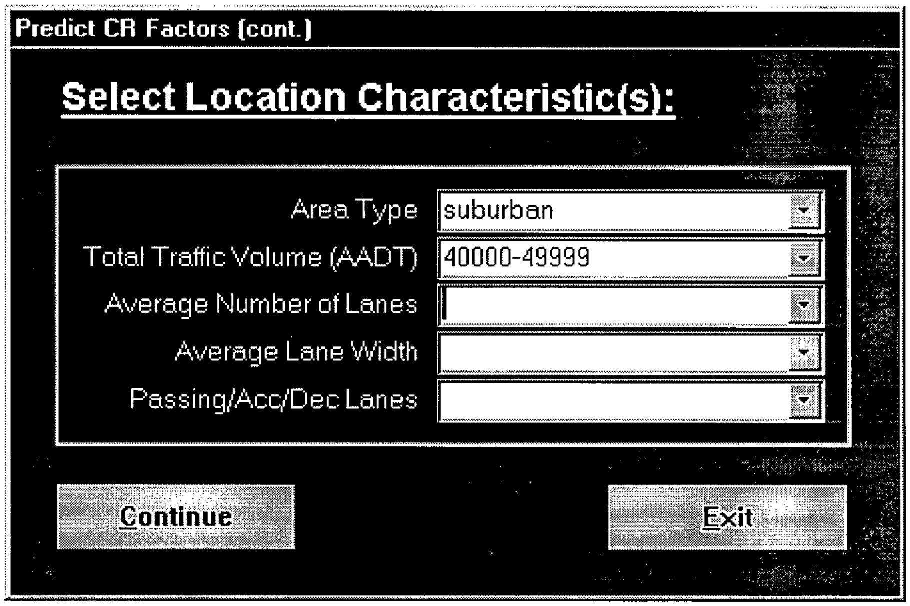


Figure D.5. ISECR Window: Input Location Characteristics for Freeways

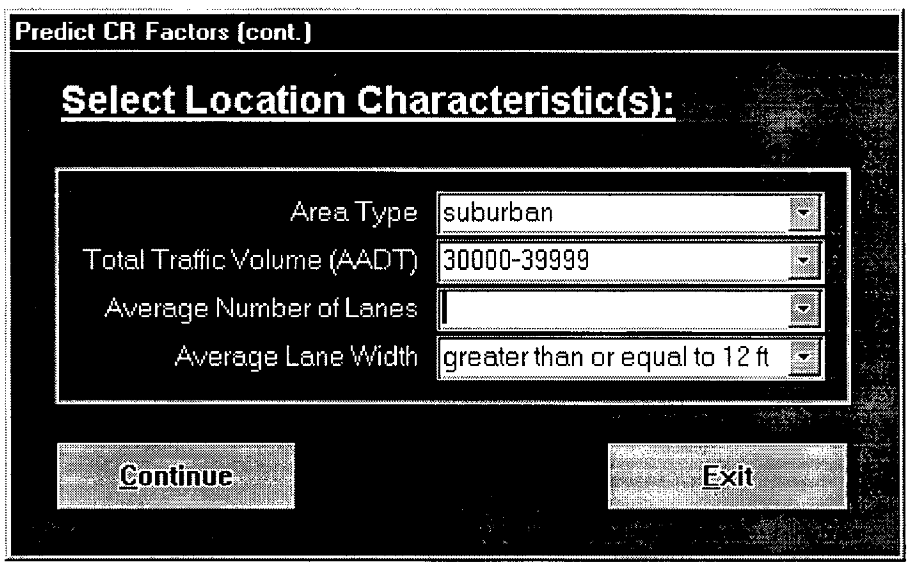


Figure D.6. ISECR Window: Input Location Characteristics for Bridges, Rails, and Construction Zones

APPENDIX E: PROBABILITY DENSITY PLOT OF BCR FOR THE PUBLIC LIGHTING EXAMPLE

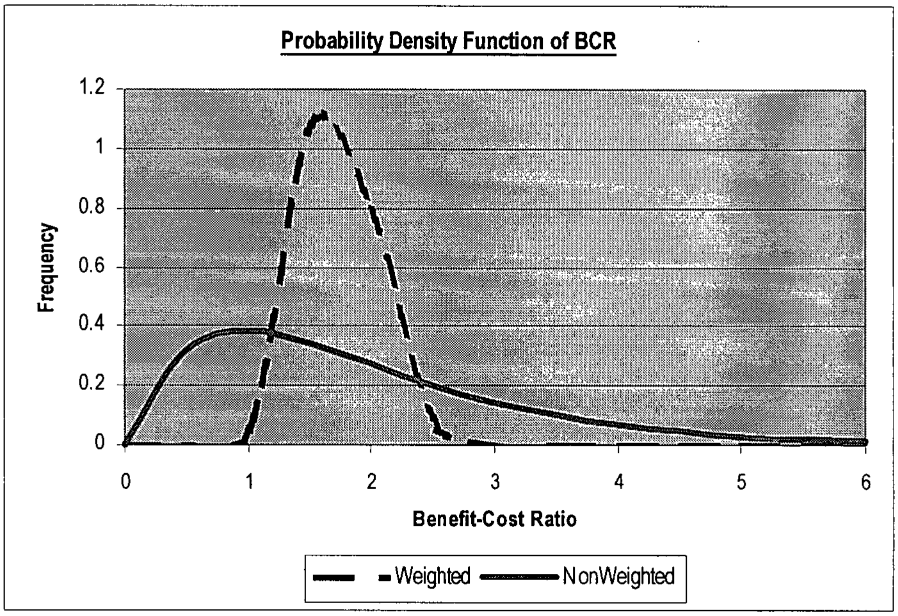


Figure E.1. Example: Probability Density Plot of BCR