Rune Elvik
TØI report 883/2007

Institute of Transport Economics
Norwegian Centre for Transport Research

State-of-the-art approaches to road accident black spot management and safety analysis of road networks

State-of-the-art approaches to road accident black spot management and safety analysis of road networks

Rune Elvik

The Institute of Transport Economics (TOI) holds the exclusive rights to the use of the entire report and its individual sections. Contents of the report may be used for referencing or as a source of information. Quotations or references must be attributed to TOI as the source with specific mention made to the author and report number. Contents must not be altered. For other use, advance permission must be provided by TOI. The report is covered by the terms and conditions specified by the Norwegian Copyright Act.

ISSN 0808-1190
ISBN 978-82-0738-8 Paper version
ISBN 978-82-0739-5 Electronic version

Oslo, September 2007

Title: State-of-the-art approaches to road accident black spot management and safety analysis of road networks

Author(s): *Rune Elvik*

TØI report 883/2007
Oslo, 2007-09
108 pages
ISBN 978-82-480-0738-8 Paper version
ISBN 978-82-480-0739-5 Electronic version
ISSN 0808-1190

Financed by:
European Commission; National Research Council of Norway

Project: 3064 RIPCORD-ISEREST

Project manager: Rune Elvik

Quality manager: Marika Kolbenstvedt

Key words:
Hazardous road locations; Accident analysis; State-of-the-art approaches; Empirical Bayes method; Network safety management

Summary:
The report describes state-of-the-art approaches (i.e. ideal approaches) to road accident black spot management and safety analysis of road networks. It is shown that the empirical Bayes approach to road safety estimation represents the best currently available approach for identifying hazardous road locations. A survey of current approaches to black spot management and safety management of road systems in different countries is made. It is argued that current approaches to accident analysis for hazardous road locations are deficient and need considerable development. Some elements of a new approach are sketched.

Tittel: Utpekning og analyse av ulykkesbelastede steder og sikkerhetsanalyse av vegsystemer

Forfatter(e): *Rune Elvik*

TØI rapport 883/2007
Oslo: 2007-09
108 sider
ISBN 978-82-480-0738-8 Papirversjon
ISBN 978-82-480-0739-5 Elektronisk versjon
ISSN 0808-1190

Finansieringskilde:
EU-kommisjonen; Norges forskningsråd

Prosjekt: 3064 RIPCORD-ISEREST

Prosjektleder: Rune Elvik

Kvalitetsansvarlig: Marika Kolbenstvedt

Emneord:
Ulykkesbelastede steder; Ulykkesanalyse; State-of-the-art tilnærmingsmåter; Empirisk Bayes metode; Sikkerhetsstyring av vegsystemer

Sammendrag:
Rapporten beskriver "state-of-the-art" tilnærmingsmåter (dvs de best tenkelige tilnærmingsmåter) til utpekning og analyse av ulykkesbelastede steder og sikkerhetsanalyse av vegsystemer. Det vises at empirisk Bayes metode for estimering av forventet ulykkestall representerer den beste metoden for å utpeke ulykkesbelastede steder. Det argumenteres for at dagens metoder for å analysere ulykker på ulykkesbelastede steder har avgjørende svakheter og bør utvikles betydelig. Enkelte elementer i en ny tilnærmingsmåte til analyse av ulykkesbelastede steder skisseres.

Language of report: English

The report can be ordered from:
Institute of Transport Economics, The library
Gaustadalleen 21, NO 0349 Oslo, Norway
Telephone +47 22 57 38 00 - www.toi.no

Rapporten kan bestilles fra:
Transportøkonomisk institutt, Biblioteket
Gaustadalleen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961
Ved gjengivelse av materiale fra publikasjonen, må fullstendig kilde oppgis

Preface

This report presents state-of-the-art approaches to two important elements of the system for safety management of roads: (1) Road accident black spot management (systems for identifying, analysing and treating hazardous road locations), and (2) Safety analysis of road networks (network safety management for the purpose of identifying and setting priorities for improving safety on longer sections of roads). The report is the first of three reports that will document work package 6 of the RIPCORD-ISEREST project (**R**oad **I**nfrastructure safety **P**rotection – **CO**re-**R**esearch and **D**evelopment for road safety in Europe; **I**n-creasing **S**af**E**ty and **RE**liability of secondary roads for a **S**ustainable surface **T**ransport).

The report is one of three reports to be published in work package 6 of RIPCORD-ISEREST. These reports are:

1. State-of-the-art approaches to road accident black spot management and safety analysis of road networks.
2. Best practice guidelines on black spot management and safety analysis of road networks.
3. Black spot management and safety analysis of road networks. Best practice guidelines and implementation steps.

Based on extensive research conducted as part of the project and on a review of similar research done by others, it is concluded that a systematic application of the empirical Bayes (EB) approach to road safety estimation represents the current state-of-the-art with respect to both black spot management and safety analysis of road networks. The empirical Bayes approach to road safety estimation has been developed by Ezra Hauer and is extensively applied in North America. It has so far not been widely applied in Europe, but was introduced successfully in Norway in 2002 when a new method for identifying hazardous road sections – injury severity density – was introduced.

The project has been funded by the European Commission and the Research Council of Norway. Rune Elvik was project manager and wrote this report. Valuable comments to earlier drafts of the report (or parts of it) have been provided by Ezra Hauer, Werner Köppel, Christian Stefan, Joao Cardoso, Rob Eenink, Martine Reurings, Stefan Matena, Roland Weber and Peter Christensen. Their comments have helped improve the report. The author is responsible for any remaining shortcomings.

Oslo, September 2007
Institute of Transport Economics

Lasse Fridstrøm
Managing director

Marika Kolbenstvedt
Head of department

Table of contents

TØI report 883/2007
Author: Rune Elvik
Oslo 2007,108 pages

Summary:

State-of-the-art approaches to road accident black spot management and safety analysis of road networks

Aim of the report and main research problems

This report presents state-of-the-art approaches to road accident black spot management and safety analysis of road networks. The term state-of-the-art approaches refers to the best approaches from a theoretical point of view, which are not necessarily identical to any approaches currently used. The report describes two activities that are essential elements of any system for safety management of roads:

1. Identification, analysis and treatment of road accident black spots (black spot management).
2. Safety analysis of road networks.

The report is part of the RIPCORD-ISEREST project funded by the European Commission. The main objective of the report is to describe state-of-the-art approaches to black spot management and safety analysis of road networks. The state-of-the-art approaches are compared to approaches that are currently used. Based on this comparison, guidelines for best practice will be developed in a subsequent report. The main research problems studied in this report are:

1. What is the state-of-the-art approach (i.e. theoretically best approach) to road accident black spot management?
2. Which approaches to road accident black spot management are currently used in European countries?
3. What is the state-of-the-art approach to safety analysis of road networks (network safety management)?
4. Which approaches to network safety management are currently used in different countries?

Elements of the state-of-the-art approach

The report concludes that a systematic use of the empirical Bayes method for road safety estimation represents the current state-of-the-art with respect to both black spot management and network safety management. The empirical Bayes approach has so far not been widely applied in Europe, but is widely used in North America.

The essential elements of an emerging state-of-the-art are as follows:

1. Black spots should be identified in terms of the expected number of accidents, not the recorded number of accidents.
2. Black spots should be identified by reference to a clearly defined population of sites, whose members can in principle be enumerated.
3. Use of a sliding window approach to identifying black spots is discouraged. This approach artificially inflates variation in accident counts.
4. To estimate the expected number of accidents, multivariate accident prediction models should be developed.
5. The best estimate of the expected number of accidents for a single site is obtained by combining the recorded number of accidents with the model estimate for that site. This should be done by applying the empirical Bayes method.
6. The performance of alternative critical values for the expected number of accidents qualifying a site as black spot should be investigated in terms of sensitivity and specificity. An optimal criterion should be chosen.
7. The traditional criterion for a true black spot, which is that there is a dominant pattern of accidents, has not been validated. Analysis of accidents at black spots is best viewed as a means of developing hypotheses regarding potentially contributing factors to the accidents.
8. Analysis of black spots should recognise the possibility that an apparent pattern may arise as a result of chance alone. Binomial tests should be applied to determine the probability that a certain number of accidents of a certain type is the result of chance only.
9. Analysis of black spots should employ a blinded design and rely on a comparison of the black spot to a safe location. The task of analysts is to identify risk factors for accidents. Analysts should not known which site is the black spot and which site is the safe one.
10. Evaluation of the effects of black spot treatment should employ the empirical Bayes before-and-after design.

A state-of-the-art approach to safety analysis of road networks should contain all these elements. In addition, a state-of-the-art approach to safety analysis of road networks should include a routine for merging adjacent sections for the purpose of accident analysis. The profiles and peaks algorithm is suitable for this purpose.

The empirical Bayes approach to road safety estimation

The empirical Bayes approach to road safety estimation has been developed by Ezra Hauer. The approach makes it possible to provide unbiased estimates of the long-term expected number of accidents for individual elements of the road system, such as a specific junction, a specific curve or a specific road section. This represents major progress in road safety estimation and permits an elimination of the bias attributable to random fluctuations in the recorded number of accidents (bias attributable to regression-to-the-mean).

Copyright © Institute of Transport Economics, 2007

By applying the empirical Bayes approach systematically, it is possible to identify hazardous road locations that have an abnormally high expected number of accidents, not just a recorded number of accidents that happened to be abnormally high due to randomness. This implies that the identification of hazardous road locations can be made substantially more accurate than before. The report presents studies comparing different criteria for identifying hazardous road locations. The empirical Bayes technique performs better than any other method according to commonly accepted epidemiological criteria of diagnostic accuracy.

There are several variants of the empirical Bayes approach. The most sophisticated version estimates the expected number of accidents by combining knowledge extracted from two sources:

1. A multivariate accident prediction model, which describes the normal level of safety and the effects of variables influencing it. The most common form of accident prediction model is a negative binomial regression model.
2. The recorded number of accidents for a specific site during the same period as that used in fitting the accident prediction model.

These two sources of knowledge are combined linearly. A weight is assigned to the normal number of accidents and a complementary weight (i.e., the weights sum to 1) is assigned to the recorded number of accidents. The better the accident prediction model (in terms of the share of systematic variation in the number of accidents explained by it), the greater is the weight attached to its predictions.

Developing good accident prediction models is difficult. The report therefore includes a review of methodological problems involved in developing and fitting accident prediction models.

Black spot management

Black spot management has a long tradition in traffic engineering and has been held in high esteem. The report describes the current approaches to black spot management in Austria, Denmark, Flanders (Belgium), Germany, Hungary, Norway, Portugal and Switzerland. The review shows that none of these countries have fully implemented a state-of-the-art approach to black spot management today. Portugal is the only country in which the empirical Bayes approach is partly implemented. There is, in general, a considerable gap between current practice and the state-of-the-art approach.

The report shows how to identify hazardous road locations by means of the empirical Bayes approach. It is moreover argued that current techniques for accident analysis need to be developed, as these techniques are not currently able to discriminate between false positives and true positives with sufficient precision.

Accident prediction models: a methodological review

Accident prediction models are important both with respect to black spot management and with respect to network safety management. The report

Copyright © Institute of Transport Economics, 2007

discusses a number of methodological problems associated with the development and fitting of accident prediction models. The discussion is concluded by a list of criteria that can be used to evaluate the quality of accident prediction models.

Network safety management

The main objective of network safety management is partly the same as that of black spot management: to identify those sites that are in greatest need of road safety treatment. There are, however, two important differences between black spot management and network safety management:

1. In network safety management, an important objective is to identify longer sections of road that have safety problems. A black spot, on the other hand, is usually a very local point on the road system, like a junction.
2. In network safety management, account is taken of accident severity and an attempt is made to identify road sections where fatal and serious injury accidents are overrepresented. In black spot management, the number of accidents at each black spot is usually too low to permit a meaningful consideration of accident severity.

Systems for network safety management in Germany, Norway and the United States are described in the report. The systems implemented in Norway and the United States are based on the empirical Bayes approach. In the United States, an algorithm designed to identify longer road sections with safety problems, profiles-and-peaks, has been implemented. If longer road sections are used, the number of accidents serving as basis for analysis is increased.

Copyright © Institute of Transport Economics, 2007

TØI-rapport 883/2007
Forfatter: Rune Elvik
Oslo 2007, 108 sider

Sammendrag:

Utpekning og analyse av ulykkesbelastede steder og sikkerhetsanalyse av vegsystemer

Rapportens formål og hovedproblemstillinger

Denne rapporten beskriver "state-of-the-art"-tilnærmingsmåter til to viktige funksjoner i et system for sikkerhetsstyring av veger:

1. Utpekning, analyse og utbedring av spesielt ulykkesbelastede steder.
2. Trafikksikkerhetsanalyse av vegnett.

Med "state-of-the-art"-tilnærmingsmåter menes de beste tilnærmingsmåter fra et teoretisk synspunkt, noe som ikke nødvendigvis sammenfaller med noen av de tilnærmingsmåter som i dag i praksis brukes for utpeke ulykkesbelastede steder eller gjennomføre sikkerhetsanalyse av vegsystemer. Rapporten er utarbeidet som ledd i EU-prosjektet RIPCORD-ISEREST. Rapportens hovedformål er å beskrive de teoretisk sett best tenkelig tilnærmingsmåter til utpekning, analyse og utbedring av spesielt ulykkesbelastede steder og trafikksikkerhetsanalyse av vegnett. Disse tilnærmingsmåter blir sammenlignet med dem som er i bruk i dag. På grunnlag av denne sammenligningen vil en senere rapport utarbeide retningslinjer for god praksis på området ("best practice guidelines"). Hovedproblemstillingene som tas opp i rapporten er:

1. Hva er den teoretisk sett beste tilnærmingsmåten til utpekning, analyse og utbedring av spesielt ulykkesbelastede steder?
2. Hvilke tilnærmingsmåter bruker ulike europeiske land til utpekning, analyse og utbedring av spesielt ulykkesbelastede steder?
3. Hva er den teoretisk sett beste tilnærmingsmåten til trafikksikkerhets-analyse av et vegnett?
4. Hvilke tilnærmingsmåter til trafikksikkerhetsanalyse av vegnett bruker ulike land i dag?

Hovedelementer i "state-of-the-art"-tilnærmingsmåter

Både når det gjelder utpekning, analyse og utbedring av spesielt ulykkesbelastede steder og når det gjelder trafikksikkerhetsanalyse av vegnett konkluderer rapporten med at den teoretisk sett beste tilnærmingsmåten er å benytte empirisk Bayes metode så systematisk som mulig. Denne metoden blir i dag i liten grad brukt i europeiske land, men er i utstrakt bruk i USA og Canada.

De viktigste elementene i de teoretisk beste metodene for utpekning og analyse av ulykkesbelastede steder og sikkerhetsanalyse av vegsystemer kan oppsummeres i følgende punkter:

1. Ulykkesbelastede steder bør identifiseres på grunnlag av forventet ulykkestall, ikke registrert ulykkestall.
2. Ulykkesbelastede steder bør utpekes med utgangspunkt i en klart definert populasjon av tilsvarende steder, som i prinsippet kan listes opp.
3. Det frarådes å benytte et "glidende vindu" til å identifisere ulykkesbelastede steder. En slik framgangsmåte forsterker problemene med falske positive.
4. For å kunne beregne forventet ulykkestall, bør det utvikles multivariate ulykkesmodeller.
5. Beste anslag på forventet ulykkestall på bestemte steder på vegnettet fremkommer ved å kombinere registrert ulykkestall med forventet ulykkestall beregnet på grunnlag av en ulykkesmodell. Dette kan gjøres med empirisk Bayes metode.
6. Ulike kritiske verdier for hvor mange ulykker som fører til at et sted regnes som ulykkesbelastet bør undersøkes med epidemiologiske kriterier og optimalt tall velges.
7. Det tradisjonelle kriteriet for å regne et sted som et ekte ulykkespunkt – at det finnes et klart ulykkesmønster – er ikke validert. Man bør kun betrakte ulykkesanalyser som en måte å utvikle hypoteser på, ikke som en test av de samme hypotesene.
8. Analyse av ulykkesbelastede steder bør ta hensyn til at tilsynelatende klare ulykkesmønstre kan oppstå rent tilfeldig. Det bør derfor gjøres statistiske tester av om ulykkesmønsteret avviker fra det normale.
9. Analyse av ulykkesbelastede steder bør skje i to trinn. Første trinn er en tradisjonell ulykkesanalyse. Andre trinn er en analyse av risikofaktorer som kan ha bidratt til ulykkene. Denne delen av analysen bør utføres som en sammenligning av et sikkert sted og et ulykkesbelastet sted. De som utfører analysen bør ikke vite hvilket sted som er sikkert og hvilket som er ulykkesbelastet.
10. Undersøkelser av virkninger av utbedring av ulykkesbelastede steder bør benytte empirisk Bayes metode.

Ved sikkerhetsanalyse av vegsystemer bør alle elementer på listen over inngå. I tillegg bør man vurdere om tilgrensende strekninger kan slås sammen for analyseformål. Til dette formål kan "profiles-and-peaks" metoden som er utviklet av Ezra Hauer benyttes.

Empirisk Bayes metode for beregning av forventingsrette ulykkestall

Empirisk Bayes metode for beregning av forventingsrette ulykkestall er utviklet av Ezra Hauer. Metoden gjør det mulig å beregne forventingsrette anslag på det langsiktige forventede ulykkestall for de enkelte elementer i vegsystemet, for eksempel ett enkelt vegkryss, en bestemt kurve eller en bestemt vegstrekning. Dette representerer et stort framskritt i metoder for estimering av trafikksikkerhet og betyr at man kan eliminere de skjevheter tilfeldige variasjoner i ulykkestall kan føre til (regresjonseffekter i ulykkestall).

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Systematisk bruk av empirisk Bayes metode gjør det mulig å utpeke spesielt ulykkesbelastede steder der det forventede ulykkestallet er unormalt høyt, i motsetning til steder der tilfeldigheter har ført til at det registrerte ulykkestallet i en viss periode var unormalt høyt. Dette innebærer at utpekningen av ulykkesbelastede steder blir langt mer treffsikker enn før. I rapporten presenteres forskning basert på norske data som viser at empirisk Bayes metode er bedre enn andre metoder som er eller har vært benyttet for å identifisere spesielt ulykkesbelastede steder.

Det finnes ulike varianter av empirisk Bayes metode. Den mest avanserte versjonen av metoden beregner forventningsrette ulykkestall ved å kombinere to kilder til kunnskap om trafikksikkerhet:

1. En multivariat ulykkesmodell som beskriver det normale nivået på trafikksikkerheten som funksjon av ulike variabler som påvirker dette. Den vanligste formen for ulykkesmodell er en negativ binomial regresjonsmodell.
2. Det registrerte ulykkestallet for det enkelte sted for den samme perioden som datagrunnlaget for ulykkesmodellen omfatter.

Disse to kildene til kunnskap vektes sammen. Jo mer av variasjonen i ulykkestall den multivariate modellen forklarer, desto mer vekt legges på dens prediksjon av ulykkestallet på det enkelte sted.

Utvikling av gode ulykkesmodeller er krevende. Rapporten inneholder derfor en drøfting av metodeproblemer knyttet til utvikling av multivariate ulykkesmodeller.

Utpekning, analyse og utbedring av spesielt ulykkesbelastede steder

Utpekning, analyse og utbedring av spesielt ulykkesbelastede steder ("black spot arbeid") har lange tradisjoner og høy status i veg- og trafikkteknikk. I rapporten gjennomgås hovedtrekkene i hvordan dette arbeidet utføres i Danmark, Flandern (Belgia), Norge, Portugal, Sveits, Tyskland, Ungarn og Østerrike. Gjennomgangen viser at ingen av landene i dag fullt ut benytter empirisk Bayes metode i sitt arbeid med utpekning, analyse og utbedring av spesielt ulykkesbelastede steder. Bare i Portugal er empirisk Bayes metode delvis tatt i bruk. Det er til dels en betydelig avstand mellom dagens praksis og den teoretisk sett beste praksis.

I rapporten vises hvordan man kan utpeke ulykkesbelastede steder med empirisk Bayes metode. Videre argumenteres det for at metodene for ulykkesanalyse på ulykkesbelastede steder må videreutvikles, fordi dagens metoder ikke skiller godt nok mellom ekte positive og falske positive ulykkesbelastede steder.

Metodologisk vurdering av ulykkesmodeller

Ulykkesmodeller spiller en viktig rolle i de teoretisk beste tilnærmingsmåtene både til utpakning av ulykkesbelastede steder og ved trafikksikkerhetsanalyse av vegnettet. Rapporten gjennomgår en lang rekke metodeproblemer knyttet til

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

utvikling av multivariate ulykkesmodeller og drøfter mulige løsninger av disse problemene. Det gis retningslinjer for å bedømme kvaliteten på ulykkesmodeller.

Trafikksikkerhetsanalyse av vegnettet

En trafikksikkerhetsanalyse av vegnettet har delvis samme siktemål som utpekning av ulykkesbelastede steder, nemlig å identifisere de deler av vegnettet som har de største ulykkesproblemene. Det er imidlertid to viktige forskjeller mellom utpekning av ulykkesbelastede steder og trafikksikkerhetsanalyse av vegnettet.

1. Ved trafikksikkerhetsanalyse av vegnettet ønsker man å finne fram til lengre vegstrekninger med ulykkesproblemer. Et spesielt ulykkesbelastet sted er derimot vanligvis et punkt på vegnettet, eksempelvis et kryss eller en kurve.
2. Ved trafikksikkerhetsanalyse av vegnettet tas det hensyn til ulykkenes alvorlighetsgrad og man ønsker å identifisere strekninger der alvorlige ulykker er overrepresentert. Ved utpekning av ulykkesbelastede steder er ofte ulykkestallet for lavt til at man på en særlig pålitelig måte kan avgjøre om alvorlige ulykker er overrepresentert.

I rapporten gjennomgås etablerte systemer for trafikksikkerhetsanalyse i Norge, Tyskland og USA. I Norge og USA bygger disse systemene på empirisk Bayes metode, i Tyskland er denne ennå ikke tatt i bruk. I USA er det utviklet en statistisk metode, ”profiles-and-peaks” (dalbunn og fjelltopper) som gjør det mulig å avgrense statistisk lengre strekninger som har en opphopning av ulykker. Ved å benytte lengre strekninger øker antall ulykker analysen kan bygge på.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

1 Introduction

This report presents state-of-the-art approaches to black spot management and safety analysis of road networks. The report is part of work package 6 of RIPCORD-ISEREST. It will be followed by a report that develops best practice guidelines for black spot management and safety analysis of road networks. The main questions discussed in this report are:

1. What are the essential elements of black spot management?
2. How are road accident black spots currently defined in different countries?
3. What is the current state-of-the art with respect to the definition, identification and analysis of road accident black spots? How do currently used definitions and techniques for identification of road accident black spots compare to the state-of-the-art?
4. What are the essential elements of road network safety management?
5. How do different countries perform safety analysis of road networks as part of network safety management today?
6. What is the current state-of-the-art with respect to safety analysis of road networks? How do current approaches to network safety management compare to the state-of-the-art?

The main objective of the report is to describe state-of-the-art techniques for black spot management and road network safety management. Current approaches will be compared to the state-of-the-art. This represents the basis for the subsequent development of best practice guidelines. Best practice guidelines will consist of a series of steps that can be taken to bring practice closer to the state-of-the-art. The report argues that a systematic use of the empirical Bayes approach to road safety estimation represents the state-of-the-art with respect both to black spot management and safety analysis of road networks. Hence, the report describes the empirical Bayes approach in some details and provides a number of examples of its use.

The report is based on an extensive literature survey, as well as on empirical studies relying on data from Norway. In addition, simulation has been used.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

2 Black spot management

This chapter presents the current approach to black spot management in some European countries. Following this presentation, elements of a state-of-the-art approach to black spot management are discussed. These elements include choice of an optimal criterion for the statistical identification of black spots, a new approach to the analysis of accidents at black spots, the development of clear criteria for discriminating between true and false black spots, an optimal approach to prioritising black spots for treatment, and an unbiased approach to the evaluation of the effects of treating black spots. Based on an empirical study, it is concluded that the empirical Bayes approach to road safety estimation represents the current state-of-the-art.

2.1 Stages of black spot management

Figure 3.1 shows the stages of black spot management in the conventional form.

Road safety management starts with the systematic collection of data that enable the identification of road safety problems, like sites that have developed into black spots. A black spot can, at this stage, be defined as any location where there is a concentration of accidents. A more precise definition will be given later on (see section 2.5). Once black spots have been identified, accidents are analysed in order to find a common pattern of accidents and factors that contribute to accidents. A visit to each site identified as a black spot is usually part of the process of analysis.

The objective of a detailed analysis of accidents and other relevant data is to identify factors contributing to accidents that may be amenable to treatment. If this analysis is not successful, it will be concluded that the black spot is likely to be false and no treatment will then be implemented. If, on the other hand, a treatment believed to be effective is found, it will be implemented and its effects evaluated.

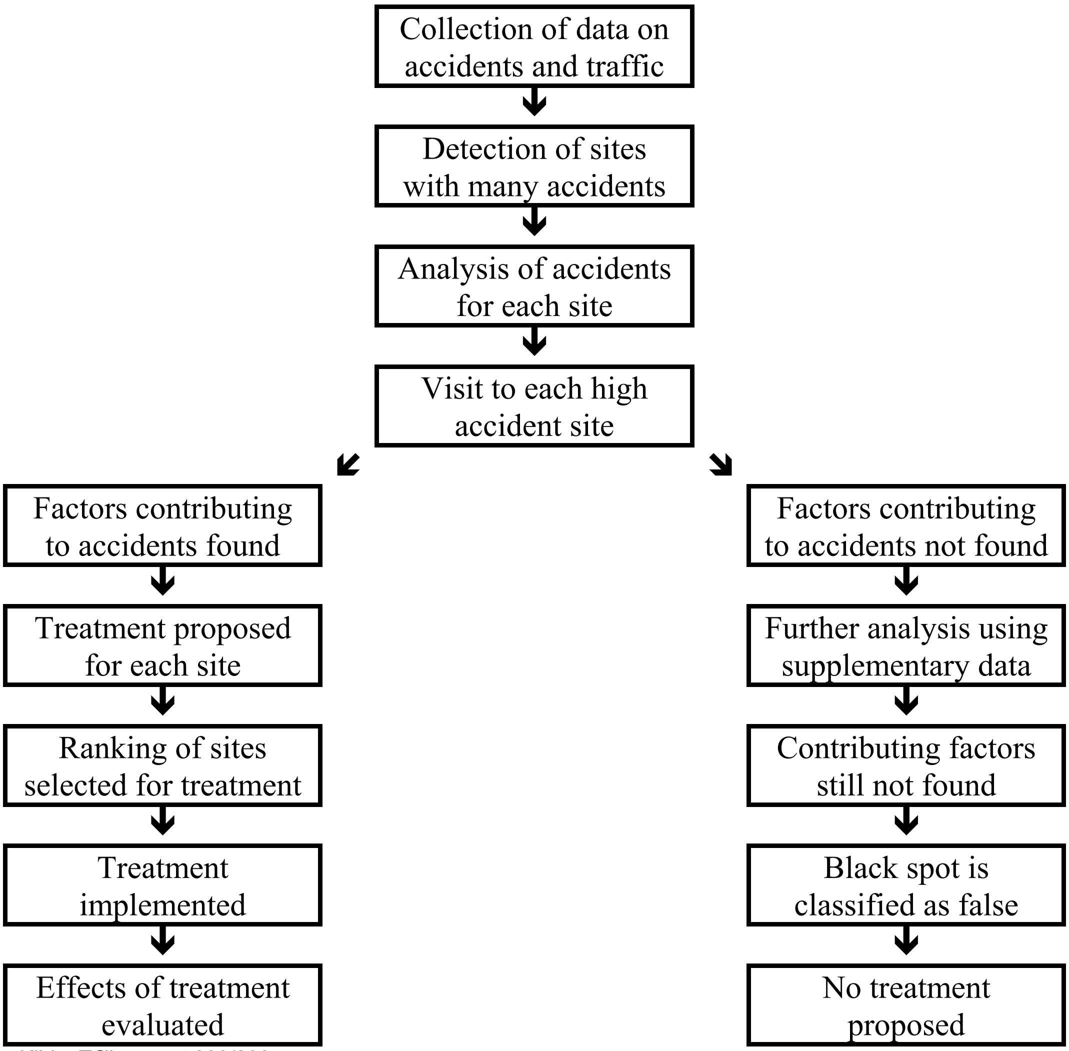
In the following sections, the elements of black spot management will be discussed more in detail, starting with the black spot concept itself.

2.2 A review of definitions of a road accident black spot

2.2.1 A taxonomy of definitions

Based on an OECD report (OECD 1976) and more recent work (Persaud, Lyon and Nguyen 1999, Hauer et. al. 2002A, Vistisen 2002, Overgaard Madsen 2005) a distinction can be made between the following common definitions of road accident black spots:

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961



Kilde: TØI rapport 883/2007

Figure 1: Stages of black spot management

1. Numerical definitions
    a. Accident number
    b. Accident rate
    c. Accident rate and number
2. Statistical definitions
    a. Critical value of accident number
    b. Critical value of accident rate
3. Model-based definitions
    a. Empirical Bayes
    b. Dispersion value

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

An example of a simple numerical definition is the official Norwegian definition of a road accident black spot: “A black spot is any location with a maximum length of 100 metres, at which at least 4 injury accidents have been recorded during the last 5 years.” This definition does not make any reference to traffic volume or to the normal number of accidents, nor does it specify the type of location considered, except by stating that the location should not extend for more than 100 metres. The identification of black spots relies on the use of a “sliding window” with a length of 100 metres.

An example of an accident rate definition of a black spot would be: “A black spot is any location (junction, section, curve, etc) at which the number of injury accidents per million vehicles (or vehicle kilometres), estimated for the most recent four years, exceeds the value of 1.5 (arbitrarily chosen)”. This definition differs from the simple accident number definition by taking account of traffic volume, and thus implicitly referring to what is regarded as a normal number of accidents.

A statistical definition of an accident black spot relies on the comparison of the recorded number of accidents to a normal number for a similar type of location. For example, a junction will be classified as a black spot if the recorded number of accidents in a specific period is significantly higher than the normal number of accidents for this type of junction. Depending on how the normal number of accidents is estimated, a statistical definition may come close to a model based definition of a black spot.

Model-based definitions of road accident black spots are derived from a multivariate accident prediction model. An example is the Empirical Bayes (EB) definition of a black spot (site with promise) given by Persaud et. al. (1999). Models were developed for intersections and road sections, and the 20 highest ranked locations were identified according to the EB estimate of the expected number of accidents.

Persaud et. al. (1999) tested the performance of two interpretations of a model-based Empirical Bayes black spot concept. According to the first definition, black spots were simply those 20 intersections that had the highest expected number of accidents, according to the EB-estimate. According to the second definition, borrowed from McGuigan (1981), a black spot was defined in terms of the potential for accident reduction, defined as follows:

Safety potential = m – P

In which m is the EB-estimate of the expected number of accidents for a specific site and P is the model estimate of the normal expected number of accidents for similar sites.

An identical concept is developed in great detail by Vistisen (2002), who refers to it as the dispersion factor. A numerical example may clarify what is meant by this. Suppose a multivariate accident prediction model has been developed, predicting 6.45 accidents ($\lambda = 6.45$) for a site. The inverse value of the over-dispersion parameter ($1/\mu$) for this model is 2.99 (i.e. the over-dispersion parameter was 0.3345). Suppose that 15 accidents have been recorded at this site ($x = 15$). The EB-estimate of the expected number of accidents is then:

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

$$E(\lambda|x) = \left(\frac{1}{1+\lambda/(1/\mu)}\right)\cdot\lambda + \left[1-\left(\frac{1}{1+\lambda/(1/\mu)}\right)\right]\cdot x \qquad (1)$$

Thus, the EB-estimate of the expected number of accidents is:

EB-estimate = [1/(1 + 6.45/2.99)] · 6.45 + [1- (1/(1 + 6.45/2.99))] · 15 = 12.29

It is now possible to decompose the contributions of three sources of variation to the recorded number of accidents:

1. General factors, included in the accident prediction model = $\frac{\lambda}{x}$ = 6.45/15 = 43 %.
2. Random variation, the excess of the recorded number of accidents to the EB-estimate = $\frac{(x - E(\lambda|x))}{x}$ = (15 – 12.29)/15 = 18 %.
3. Local factors (and unknown or unmeasured general factors), the difference between the EB-estimate and the model estimate = $\frac{(E(\lambda|x) - \lambda)}{x}$ = (12.29 – 6.45)/15 = 39 %.

It is the third of these factors that Vistisen (2002) refers to as the dispersion factor. Ideally speaking, it represents the effects of local risk factors that are not included in the accident prediction model. In practice, however, the term is likely to include both the effects of local factors and the effects of more general factors that have not been included in the accident prediction model.

Some examples from European countries show the multiplicity of definitions of road accident black spots currently used.

2.2.2 Definition, identification and analysis of black spots in Austria

In Austria, black spots are defined in the Austrian Guideline Code for the Planning, Construction and Maintenance of Roads (RVS 1.21) published in November 2002. According to this guideline, scenes of accidents are distinguished in black spots and hazardous locations, depending on their recorded crash history. To be classified as a black spot, one of the following two criteria has to be met:

- 3 or more similar injury accidents within 3 years and a relative coefficient R_k of at least 0.8. The value of this coefficient is calculated as follows:

$$R_K = \frac{U}{0.5 + 7 \text{ x } 10^{-5} \text{ x AADT}}$$

Where:

AADT = Annual Average Daily Traffic [vehicles/24 hours]

U = Number of injury accidents within 3 years

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

- at least 5 accidents (including property damage only) of similar type within one year. Since 1995 property damage accidents are not recorded in Austria, hence black spot management primarily relies on the first definition.

For calculation, a sliding window with a length of 250 m is being used. The window follows the course of the road (network) under surveillance and flags each location where one of the two criteria for a black spot is met (see Figure 2).

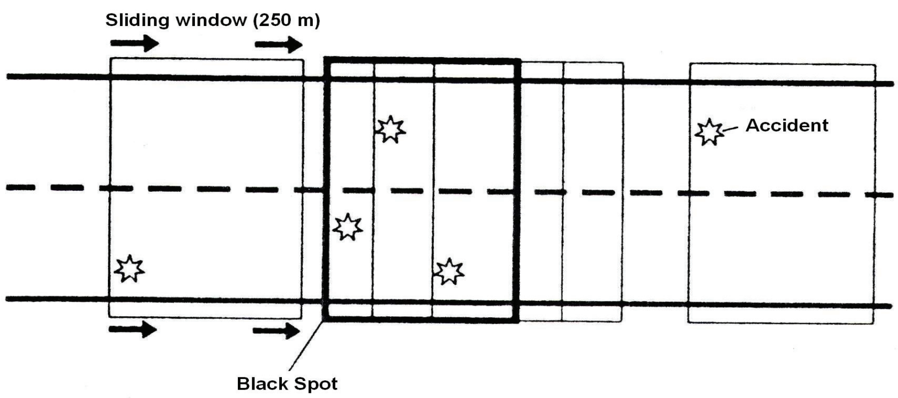


Figure 2: Identification of road accident black spots in Austria by sliding window approach. Source: Austrian guidelines for black spot identification.

The critical value of 0.8 of the relative coefficient R_k will be reached under the following circumstances:

3 injury accidents in 3 years and an AADT up to 10,700 vehicles/24 hours

4 injury accidents in 3 years and an AADT up to 16,700 vehicles/24 hours

5 injury accidents in 3 years and an AADT up to 22,600 vehicles/24 hours

6 injury accidents in 3 years and an AADT up to 28,600 vehicles/24 hours

Figure 3 shows the different threshold levels for the coefficient R_k, depending on the number of (similar) injury accidents and AADT. The grey box on the bottom of the graph marks the non-critical range of the accident analysis with an R_k below 0.8. If there is no data available on annual average daily traffic, the location is classified as a possible black spot according to the first criterion listed (3 or more similar injury accidents in 3 years).

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

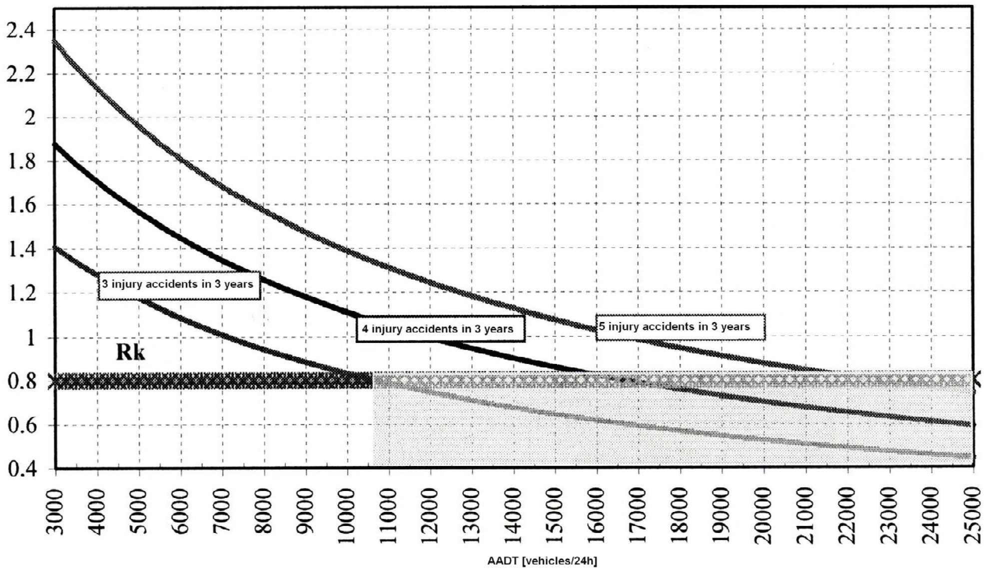


Figure 3: Critical values of the coefficient Rk for identification of road accident black spots in Austria. Source: Austrian guidelines for black spot identification.

The Austrian Guideline Code for the Planning, Construction and Maintenance of Roads classifies accidents according to how they occur and which road users are involved. Accidents have to be of similar type before one of the two criteria defining a black spot can be applied. A distinction is made between the following types of accident:

- Single vehicle accident
- Rear end collision
- Head-on collision
- Right angle collision
- Collision involving parked vehicle
- Pedestrian accident
- Cyclist accident
- Accident involving powered two wheeler
- Accident in twilight or darkness
- Accident on wet road surface or on road surface covered by ice or snow

It can be seen that these categories are not mutually exclusive, i.e. the same accident may be placed in more than one category (e.g. a pedestrian accident in the dark on a wet road surface).

As mentioned previously, accidents with property damage only have not been systematically recorded since 1995 and therefore can not be used for locating black spots. Hence, black spot Management in Austria currently relies on injury accidents only. If such an accident occurs on a public road, the police force is obliged to fill out a standardized accident report form. This information is delivered to both the Statistics Austria and the Austria Road Safety Board (KfV) for further investigations and analyses.

Although there is a legal basis for black spot management in Austria, the definition of black spots described in the RVS 1.21 is not binding, i.e. each of the

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

nine federal states is free to make changes or define a black spot by themselves and thus influence their amount on their respective road network. Some of the states carry out in-depth investigations and propose certain measures to prevent further accidents while others do not bother at all with the accident data the Austrian Road Safety Board provides them with.

Black spot management in Austria consists of the following steps:

A) Statistical analysis of the black spot

- Type of accident (Rear-end collision, frontal collision, etc.)
- Weather and Road condition during the accident
- Road user involvement (lorry, passenger car, pedestrian, etc.)
- Severity of casualty (fatal, seriously or slightly injured, unharmed)

B) Local Assessment of the black spot

In-depth investigation of the accident site (including evaluation of the street environment, light conditions during darkness and twilight, traffic guidance, etc.)

C) Proposing measures

On the basis of the analysis and local assessment of the black spot, remedial measures are proposed. Some typical treatments for both black spots and hazardous locations are set out below. The treatment(s) selected should be those that address the safety problem in the most cost-effective manner.

- Road realignments
- Sealing of road shoulders
- Speed control measures (e.g. radar boxes)
- Installation/upgrading of street lighting
- Installation of pedestrian signals
- etc.

D) Implementation of measures

Measures are implemented to the extent available financial funds makes it possible.

E) Before-and-after comparison of accidents occurrence

After implementation of measures, accident occurrence is being observed in order to evaluate if the number and severity of accidents has been reduced. If this is not the case, further action has to be considered.

2.2.3 Definition of black spots in Denmark

The definition of road accident black spots in Denmark relies on a fairly detailed classification of the road system into various types of road sections and various types of intersections (Vistisen 2002, Overgaard Madsen 2005). For national roads, a distinction is made between road sections, roundabouts and other intersections. Each of these groups in turn consists of several types, such as motorways, other dual carriageway roads, two-lane roads in rural districts, two-lane roads in urban areas, and so on. In each group of road sections, the normal

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

expected number of accidents is estimated by applying the following simple prediction model:

$$\text{Normal number of accidents} = \boldsymbol{\alpha} \cdot \boldsymbol{AADT}^{\boldsymbol{\beta}} \qquad (2)$$

AADT is annual average daily traffic volume. For intersections, models of the following form are used:

$$\text{Normal number of accidents} = \boldsymbol{\alpha} \cdot \boldsymbol{AADT}_{\boldsymbol{ma}}^{\boldsymbol{\beta}1} \cdot \boldsymbol{AADT}_{\boldsymbol{mi}}^{\boldsymbol{b}2} \qquad (3)$$

The subscripts *ma* and *mi* refer to the major and minor approach to an intersection. In fitting these models, data for a period of 3 to 5 years are used.

To identify a black spot, a test relying on the Poisson distribution is used. It is assumed that the models fitted adequately represent systematic variation in the number of accidents to the extent this variation can be attributed to traffic volume and the variables used to classify roads and intersections. If a site has a significantly higher number of accidents than random variation alone can explain, it is therefore assumed that the excessive number of accidents must be at least partly due to local risk factors. Hence, the definition of a road accident black spot used in Denmark is: a site with a reported number of accidents, which is higher than both a fixed minimum number and significantly higher than the normal expected number of accidents for a similar type of roadway element (section, roundabout or intersection).

Currently, the minimum number of accidents for a site to be identified as a black spot is 4 accidents recorded during a period of 5 years. Moreover, the level of significance used in the statistical test is 5%. Thus, suppose the normal number of accidents for a roundabout has been estimated to 2.8 (during 5 years) and that 5 accidents have been recorded. Applying the Poisson distribution, the probability of observing at least 5 accidents given that the mean number is 2.8 can be calculated to 0.152, which means that this roundabout would not be classified as a black spot.

As far as road sections are concerned, black spots are identified by means of a sliding window approach, similar to the one used in Austria (see section 2.2.2). In Denmark, however, the length of the window used is found by solving the following equation with respect to l (Vistisen 2002, page 125):

$$p(X \geq x_{min} | \lambda_i T_i) = 1 - \sum_{x=0}^{x\,min-1} \frac{(\lambda_i \cdot T_i \cdot l)^x}{x!} \cdot e^{-(\lambda_i \cdot T_i \cdot l)} = \alpha \qquad (4)$$

Here, the letter α refers to the critical level of statistical significance, i.e. 5%. X is the recorded number of accidents and $p(X \geq x_{min} | \lambda_i T_i)$ is the probability that the recorded number of accidents exceeds the minimum value for a black spot, given that the expected number is $\lambda_i T_i$. T denotes time, l denotes road length and λ denotes the expected number of accidents per kilometre of road. The equation

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

implies that a shorter sliding window will be used for road sections that have a high normal number of accidents than for road sections that have a low expected number of accidents. The value of the length of the sliding window will be determined so that for this length, the probability of observing an accident number equal to or greater than the minimum critical value (i.e. 4 accidents) equals 5%.

As an example, consider a road that has an average accident density of 1.5 accidents per kilometre. For such a road, the probability of observing 4 or more accidents is 0.0656, i.e. more than the critical probability value of 0.05. In general, the probability of observing a count of 4 or more is equal to 0.050 in a Poisson distribution whose expected value is 1.366. Hence, the length of the sliding window should be fixed to a value for which the mean number of accidents captured within the window is equal to 1.366. Thus, if the mean number of accidents per kilometre is 1.5, each sliding window should have a length of 0.909 kilometres, as the mean per section (if we imagine the sections were fixed and not sliding) then becomes 1.366.

If a road section had 10 accidents per kilometre, each sliding window would have to be 0.137 kilometres, at the expected number of accidents per 0.137 kilometre section is 1.366.

The Danish model for identifying black spots comes close to a model-based approach. The main difference between the Danish approach and a model-based approach is that black spots in Denmark are identified according to the recorded number of accidents, not the expected number.

Following accident analysis, potential safety treatments are prioritised on the basis of their estimated first year rate of return. The first year rate of return corresponds to the value of accident costs saved during the first year after treatment divided by the costs of implementing the safety treatment. Sites are ranked for treatment according to the marginal first year rates of return. Marginal first year rates of return are determined by ranking alternative treatments for a single site, or alternatives sites for treatment according to the following criterion:

$$\boldsymbol{MB}_{i,Y1Y2} = \frac{\boldsymbol{AC}_{iY2} - \boldsymbol{AC}_{iY1}}{\boldsymbol{C}_{iY2} - \boldsymbol{C}_{iY1}} \qquad (5)$$

AC represents the savings in accident costs, C represents the cost of implementing a measure, and 1 and 2 are alternative measures, for which the first year rate of return is highest for measure 2.

2.2.4 Definition and analysis of road accident black spots in Flanders

In Flanders, the following definition of a road accident black spot is applied (Geurts 2006), based on police reports of accidents:

Each site where in the last three years three or more accidents have occurred, is selected. Then, a site is considered to be dangerous when its score for priority (S), calculated using the following formula, equals 15 or more:

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

S = LI + 3SI + 5DI

where LI = total number of slight injuries

SI = total number of serious injuries (Each casualty that is admitted more than 24 hours in hospital)

DI = total number of deadly injuries (Each casualty that died within 30 days after the accident)

Based on this definition, in Flanders currently 1,014 locations are considered as black spots. Each location identified as a black spot should have a length of not more than 100 metres. Three years of accident data are used to identify black spots.

Unlike the other definitions of a road accident black spot reviewed so far, the definition used in Flanders tries to account for accident severity by assigning a greater weight to serious and fatal injuries than to slight injuries. Thus, a site at which there has been 2 fatal accidents and 1 serious injury accident will get a priority score of 13, whereas a site at which there has been 10 slight injury accidents will get a priority score of 10.

Geurts (2006) investigated the sensitivity of the ranking of black spots in Flanders to the choice of weights for injury severity. Not surprisingly, the choice of weights was found to influence the ranking of the black spots.

Geurts (2006) also explored the use of data mining to analyse accidents at black spots. Systematic data mining can be viewed as a technique for extracting information on all potentially interesting patterns in the accidents recorded at a black spot. It should not, however, be seen as a replacement for site visits or other techniques of analysis that go more in depth.

2.2.5 Definition and analysis of black spots in Germany

In Germany, a distinction is made between black spots, black sections and black areas. The distinction between these types of accident concentrations is made by examining accident maps.

A black spot is defined as follows: A site is considered a “frequent-accident spot” if a large number of accidents occur on a very small section of a road in a road network, i.e. if a certain number of accidents is reached or exceeded on the one-year and/or three-year map. Typical frequent-accident spots may include intersections, road/road and road/off-road track junctions, bends, humps, railway crossings and inclines.

A preliminary investigation of a frequent accident spot should be conducted if one of the limit values shown in Table 1 is reached or exceeded. The limit values suggested in Table 1 apply to the road network both within and outside built-up areas as well as motorways.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 1: Critical values for identification of black spots in Germany. Source: German road and transport research association, 2006

Source of data	Critical count of accidents	Length of period
1-year map	5 of similar type	12 months
3-year map	5 injury accidents	36 months
3-year map	3 serious injury accidents	36 months

Kilde: TØI rapport 883/2007

To help identify sites that have a concentration of similar types of accident, a distinction is made between the following types of accident:

- Single vehicle accident (driving accident)
- Turning-off accidents
- Turning-into/crossing accidents
- Pedestrian crossing accident
- Accident involved stopped or parked vehicle
- Accidents involving longitudinal traffic (rear-end and head-on collisions)
- Other accident

The "other accident" category is used for any accident that does not fit any of the other categories. Furthermore, a distinction is made between the following levels of accident severity:

- Fatal accident (at least one person killed within 30 days)
- Accident in which at least 1 person is seriously injured
- Accident in which at least 1 person is slightly injured
- Accident with severe material damage
- Other material damage accident

Accidents are classified according to the most serious injury occurring in the accident. Accident severity is taken into account in identifying black spots, in that the critical values are lower for injury accidents than for all accidents (irrespective of severity), and lower for serious injury accidents than for all injury accidents.

"Frequent-accident lines" (black sections) (FAL) are accident concentrations along lengthy sections of road. They are examined in more detail if they occur on the (3-YM) accident-type map of accidents with serious personal injury.

"Frequent-accident areas" (black areas) (FAA) mostly occur in built-up areas (relatively large municipalities and in towns) and usually within the residential road network. Accident occurrence in (residential) areas is primarily assessed on the basis of the three-year map showing the accidents with personal injury.

Once black spots, black sections and black areas have been identified by means of accident maps, a preliminary investigation is performed. The preliminary investigation considers three aspects: (1) The impact area of a black spot, section or area, that is how far does it extend geographically. The geographic extension is determined according to the accident map and is then used in later stages of analysis. (2) The length of the period of analysis. Periods of 1, 2 or 3 years are used most often. A longer period can be used if there is reason to believe that there is a temporal trend in accident occurrence. (3) For black spots that have a

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

large number of accidents (clearly exceeding the minimum values), the trend over time is investigated. The main objective is to detect whether the site shows signs of improving or whether it is getting worse. This analysis of trends over time is used rather than setting a stricter critical value for sites that have large traffic volumes.

Preliminary analysis is followed by a detailed in-depth accident analysis designed to identify why the accidents occur and propose treatments. Sites are ranked for treatment according to the following criteria. Simple black spots (those that fulfil the first criterion listed in Table 1) are ranked according to the number of comparable accidents. Black spots identified according to the number of injury accidents, are ranked according to the number of serious injury accidents. In case of ties, the total number of injury accidents is used. For black spots that have a large number of accidents, two ranking criteria are used. One of the criteria is based on accident severity, which is summarised in terms of the total accident costs. The other criterion is based on the number of similar accidents.

Black sections are ranked for treatment according to the density of serious injury accidents, i.e. the number of serious injury accidents per kilometre of road per year. This criterion is also applied to black areas.

2.2.6 Definition of black spots in Hungary

In Hungary, two definitions of road accident black spo6ts are used. Outside built-up areas, a black spot is defined as a location where at least 4 accidents have been recorded during 3 years on a road section no longer than 1000 metres. Inside built-up areas, a black spot is defined as a location where at least 4 accidents have been recorded in 3 years on a road section no longer than 100 metres.

Search for black spots with scaled accident sites map or data lists is made by using the so-called "sliding window" approach. The „window" is either 1000 metres or 100 metres wide.

Once black spots have been identified, they are ranked for further study. The aim is to perform detailed engineering studies for about 10-15 % of the black spots that have been identified statistically. When ranking black spots, traffic volume is taken into account, so as to identify those black spots that have a higher then normal accident rate. Thus, the Hungarian approach to the definition of black spots in practice comes close to a rate and number method.

2.2.7 Definition and analysis of black spots in Norway

In Norway, a distinction is made between black spots and black sections. A black spot is any location with a length of not more than 100 metres where at least 4 injury accidents have been recorded in the last 5 years. A black section is any road section with a length of not more than 1000 metres where at least 10 injury accidents have been recorded during the last 5 years. The period used to identify black spots or black sections was recently extended from 4 to 5 years (Statens vegvesen, håndbok 115, 2006, draft version).

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Black spots and black sections are identified by applying a sliding window, which is fitted to the location of the accidents. Black sections will often consist of several black spots that are located near one another.

Once black spots or black sections have been identified, they are ranked for further detailed study. Ranking consists of performing the following four steps:

1. Estimate the cost of accidents based on the recorded number of accidents.
2. Estimate the expected number of accidents and the cost of these accidents for a similar spot or section that has the best possible road safety standard.
3. Estimate the probability that the recorded number of accidents exceeds the number that can expected at a similar site that has the best possible road safety standard.
4. Rank sites (spots or sections) according to the difference between the cost of accidents estimated for the site and the cost of accident for a similar site that has the best possible road safety standard (the potential for reduction of accident costs).

According to the Norwegian guidelines, the ranking of sites by their potential for safety improvement corresponds to the ranking criterion used in Germany in road network safety management (see also chapter 4).

Step 1 of the ranking analysis is based on the recorded number of accidents and on mean values for accident costs. The mean values have been estimated on the basis of a large number of accidents in order not to be greatly influenced by random fluctuations.

Step 2 is based on a comprehensive set of normal accident rates for various roadway elements. Accident rates are stated as the number of injury accidents per million entering vehicles or per million vehicle kilometres. The best possible safety standard has been defined as the expected cost of accidents at an accident rate 20% below the mean accident rate for a given roadway element.

To evaluate the probability that the recorded number of accidents exceeds the normal expected number of accidents for a similar site, a nomogram has been developed showing probability values corresponding to the 20%, 10%, 5% and 1% levels of statistical significance. These values were determined as the critical percentiles of the Poisson distribution.

Finally, in step 4, sites are ranked according to the size of their potential for safety improvement, stated in terms of the reduction of accident costs that can be attained if the best possible safety standard is achieved. This ranking is performed only for sites that have an abnormally high number of accidents according to the statistical test (step 3).

Sites that are highly ranked are selected for a more detailed engineering study. This includes a detailed accident analysis, site visits, observations of road user behaviour, etc. Based on the detailed analysis, measures to improve safety are proposed and the costs and effects of these are estimated. Priorities for implementation of treatments are set according to the net benefit-cost ratio estimated for each site.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

2.2.8 Definition of black spots in Portugal

In Portugal, black spot detection is carried out for roads belonging to the National Road Network, which is managed by the Portugese Highways Agency (*Estradas de Portugal* - EP).

Two definitions of black spot are currently used by EP: one was set by the Traffic Directorate (*Direcção-Geral de Viação* - DGV); the other was proposed by LNEC. According to the definition by DGV, an accident black spot is a road section with a maximum length of 200 m, with 5 or more accidents and a severity indicator greater than 20, in the year of analysis. No distinction is made between intersection and non-intersection accidents. The total number of accidents is used. The severity index is calculated by the following weighted sum:

100 · number of fatalities + 10 · number of serious injuries + Number of slight injuries.

Detection is carried out using a sliding window moving along the road. The results are published in the yearly report of the Portuguese Road Safety Observatory, a DGV branch.

An alternative method was proposed by LNEC in 1997 and tested in 1998. Currently it is being applied as part of Portuguese Road Safety Plan 2004-2010. In this method, a theoretical definition similar to the one presented in section 2.5 is adopted: a black spot is a geographical area where the expected accident frequency is greater than in similar (not necessarily adjacent) areas, due to the influence of road characteristics peculiar to the area. In practice this definition is applied differently to intersection and non-intersection accidents.

As far as non-intersection accidents are concerned, different minimum road section lengths are used for single carriageway roads and in dual carriageway roads: 250 metres minimum length is used in the first case, and 500 metres in the second. In dual carriageway roads the detection is carried out separately in each carriageway.

The road network was divided in 6 classes of road, according to the number of carriageways (single or dual), carriageway width of single carriageway roads (below or equal to 6.00 m, not greater than 7.00 m, below or equal to 7.75 m, and greater than 7.75 m) and number of lanes in each separate carriageway (two lanes and three or more lanes). For each class of road, a unique accident prediction model is fitted to accident data for a five year reference period. Initially, models fitted to 1994-1998 data were applied; currently models fitted to 1999-2003 data are being used. Models were developed using the following general equation:

$$E(\lambda) = \beta 1 \times ADT^{\beta 2} \times CW^{\beta 3} \times e^{(\beta 4 \times ADT/1000)} \quad (6)$$

Where:

$E(\lambda)$ = expected number of accidents in the five year period.

ADT = Average daily traffic.

CW = Carriageway width.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

β_i = Parameters estimated with model fitting.

In most cases, only the average daily traffic was kept as a statistically significant explanatory variable.

In each year, observed accident data from the previous five years are combined with the corresponding accident prediction model to estimate the expected number of accidents in each 250 m single carriageway (or 500 m dual carriageway) road section of each road class. In practice, this is an application of the empirical Bayes method, as described by Hauer et al. (2002B). Depending on the road class, the worst 1/1000 or 2/1000 cases are selected for detailed safety diagnosis and possible intervention.

Intersections are treated separately. As no applicable accident models have been developed yet, the worst 20 intersections in each road class are selected for detailed accident analysis.

2.2.9 Definition of black spots in Switzerland

In Switzerland a definition of Black Spot Analysis and treatment can be found in the standards of the Swiss Association of Road and Transport Experts. Regrettably the application of these tools is not institutionalized. However, some Departments of Transportation as well as some Cantonal Polices apply it. Institutions like the Swiss Council Of Accident Prevention or the Institute For Transport Planning and Systems of The Swiss Federal Institute Of Technology Zurich apply it for counselling.

Basically a black spot is defined as a section of a road or an intersection, where the number of accidents is “well above” the number of accidents at comparable sites. Comparable sites are identified by the following procedure:

1. The section to be analysed is divided into two different types of road: *Open road intervals* and *intersection intervals*. *Open road intervals* have a regular length that depends on the type of road. The more important the road the longer are the *open road intervals* (varying between 100m and 500m). *Intersection intervals* contain major intersections.
2. An *average open road accident rate* for the whole section is calculated based on the accidents that occur on *open road intervals.*
3. An *average intersection accident rate* for the whole section is calculated based on accidents that occur on *intersection intervals*.
4. For both the *open road accident rate* and the *intersection accident rate* a range is calculated. The calculated accident rates should fit within this range (which is basically the confidence interval with a probability of error of 5%). This range represents the expected accident rate calculated on the basis of the analysed intervals.
5. So called *accident sites* are determined. *Accident sites* are *open road intervals* or *intersection intervals* where the number of accidents exceeds the estimated expected number of accidents based on normal accident rates.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

6. On the last step only the *accident sites* are compared to so called *threshold values*. The *accident sites* with accident numbers exceeding the *threshold values* are considered to be *black spots*. The *threshold values* vary depending on the road type (Table 2) and refer to a period of 2 years.

The identified black spots are redesigned based on a specific procedure, the so called *technical accident analysis*. The *technical accident analysis* is based upon two separate approaches. These analyses should be conducted by two independent engineers.

1. The infrastructure analysis. It consists in the comparison of the site with the standards for the purpose of determining deficiencies. These deficiencies are called *detected deficiencies*.
2. The analysis of the determinant accidents. It aims at determining probable deficiencies that may lead to this type of accident. These deficiencies are called *probable deficiencies.*

Table 2: Threshold values to determine black spots

Road type and location	Threshold values		
	Number of all crashes	**Number of crashes with injuries and/or fatalities**	**Number of crashes with fatalities**
Motorways (per direction)	10	4	2
Entrance/exit ramps	10	4	2
Rural roads (intersections and open roads)	8	4	2
Main roads in built-up areas - open roads	8	5	2
Main roads in built-up areas - intersections	10	6	2

Kilde: TØI report 883/2007

The comparison of the *detected deficiencies* with the *probable deficiencies* leads to the *determinant deficiencies*. These *determinant deficiencies* are to be corrected. This insight implies that not all *detected deficiencies* lead to accidents. This procedure is intended to minimise the bias of judging a location knowing the accident types and vice versa. Moreover it allows a more efficient use of resources.

2.2.10 Comparative analysis of methods for identifying road accident black spots

The various definitions of road accident black spots and the techniques used for identifying them reviewed above differ along the following dimensions:

1. Whether they make a reference to a population of sites (a specific type of roadway element or just any location) or not.
2. Whether black spots are identified by means of a sliding window or by reference to a set of given locations.
3. Whether they make a reference to the normal level of safety or not.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

4. Whether they are based on the recorded number of accidents or an estimate of the expected number of accidents.
5. Whether accident severity is considered or not.
6. The length of the identification period used.

Table 3 provides an overview of the definitions presented above in terms of these dimensions. It is seen that none of the definitions listed are identical in all respects.

Table 3: Overview of definitions of road accident black spots in selected European countries

Country	Reference to population of sites	Sliding window applied	Reference to normal level of safety	Recorded or expected number of accidents	Accident severity considered	Length of identific-ation period
Austria	No	Yes, 250m	Yes, by means of critical values for accident rate	Recorded, minimum critical value 3 – function of traffic volume	No	3 years
Denmark	Yes, detailed categoris-ation of roadway elements	Yes, for road sections – variable length	Yes, by means of accident prediction models	Recorded, based on statistical test – minimum 4 accidents	No	5 years
Flanders	No	Yes, 100m	No	Recorded, weighted by severity	Yes, by means of weights	3 years
Germany	No	No, accident maps inspected	No	Recorded, minimum values 3 or 5	Yes, by different critical values	1 year (all accidents) or 3 years (injury accidents)
Hungary	No	Yes, 100m or 1000m	No	Recorded, minimum 4	No	3 years
Norway	Not when identifying black spots	Yes, 100m (spot) or 1000m (section)	Yes, by means of normal accident rates for roadway elements	Recorded higher than normal by statistical test, mini-mum values 4 (spots) or 10 (sections)	Yes, by estimating accident costs and potential savings	5 years
Portugal	Yes, for one definition; no for the other	Yes, for one definition; no for the other	Yes, for one definition; no for the other	Recorded in one definition (minimum 5), expected in the other	Yes in one definition (by severity weighting), no in other	1 year or 5 years
Switzerland	Yes, open roads and junctions	No, fixed sections of variable length	Yes	Recorded, a set of critical values	Yes, by different critical values	2 years

Kilde: TØI report 883/2007

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Black spots are, in most countries, not identified by sampling units from lists of elements belonging to a population of sites, all members of which are in principle enumerable, such as “all three legged junctions”, “all horizontal curves with a radius less than 200 metres”, or similar. In most countries, black spots are identified by applying a sliding window to the locations of accidents, and fixing the position of the window at points where it contains the (local) maximum number of accidents. In Germany accident maps are used, but in practice this may come to nearly the same thing as using a sliding window, since black spots are identified according to the locations of accidents. Denmark uses a sliding window for road sections, but not for junctions. Portugal uses a sliding window for one of its definitions of an accident black spot, not for the other. Switzerland does not use a sliding window.

An accident black spot is generally taken to be a site that has an abnormally high number of accidents. This definition suggests that black spots cannot be meaningfully identified without some reference to the normal level of safety. Some of the currently employed definitions of black spots in European countries make an explicit reference to the normal level of safety, but – surprisingly – not all definitions make such a reference. References to the normal level of safety are generally made by comparing the number of accidents at sites identified as black spots to the number of accidents expected for similar sites, estimated by means of accident prediction models or by referring to a set of normal accident rates.

Black spots are in all countries identified in terms of the recorded number of accidents. The only exception from this is the black spot definition develop by LNEC in Portugal, which relies on the empirical Bays method. Defining black spots in terms of the recorded number of accidents is perhaps not very surprising, as the long-term expected number of accidents cannot be observed, only estimated. In some countries, tests are performed to determine if the recorded number of accidents is significantly higher than the normal number expected for similar sites. Presumably, sites that do not pass this test (i.e. the test does not show a statistically significant difference in the number of accidents: recorded versus normal) are deleted from the list of black spots and not treated as abnormal.

Some definitions of black spots consider accident severity, other definitions do not. If accident severity is considered, there is no standard way of doing so. Three different approaches can be identified. One approach is to set a more stringent critical value for the number of serious injury accident accidents than for all injury accidents when identifying black spots. A second approach is apply weights to accidents at different levels of severity. A third approach is to estimate the costs of accidents. These costs vary according to injury severity; hence, costs will be higher at sites that have a high proportion of fatal or serious injury accidents.

The length of the period used to identify black spots varies from 1 year to 5 years. A period of 3 years is used frequently. Research by Cheng and Washington (2005) shows that the gain in the accuracy of black spot identification obtained by using a longer period of 3 years is marginal and declines rapidly as the length of the period is increased. There is little point in using a longer period than 5 years.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 4 illustrates how varying the length of the identification period influences the mean number of accidents observed during a second period of identical length for identification periods of 1, 2, 3 or 4 years. The table is based on Norwegian data referring to road sections of 1 kilometre on national roads. The road sections were fixed. No sliding window was applied.

As can be seen from Table 4, if a period of only 1 year is used, there is a considerable regression-to-the-mean effect for all counts of accidents from 0 to 10. For 6 out of the 11 different counts of accidents included in Table 4 (counts of 0, 1, 2 etc) this effect is reduced when the length of both periods is extended to 2 years. Lengthening the periods to 3 years further reduces regression-to-the-mean (compared to 2+2 years) for 7 out of 11 different counts of accidents. Lengthening from 3+3 to 4+4 years, however, reduces the regression-to-the-mean effect for only 2 of the 11 different counts of accidents included in Table 4.

Table 4: Effects of varying the length of the period of identification on the stability of accidents counts in a subsequent period of identical length

	Mean number of accidents during "second" period for sites that had 0,1,2 etc accidents during the "first" period for periods of different length			
Count of accidents in first period	**1 year + 1 year**	**2 years + 2 years**	**3 years + 3 years**	**4 years + 4 years**
0	0.099	0.166	0.172	0.137
1	0.349	0.495	0.443	0.404
2	0.834	0.936	1.053	0.771
3	1.404	1.541	1.616	1.465
4	2.207	2.054	2.455	2.281
5	3.500	3.606	3.327	2.020
6	4.778	4.536	3.448	2.935
7	3.556	5.167	5.750	4.154
8	7.375	5.214	5.750	3.000
9	8.333	5.500	6.667	7.143
10	3.333	7.700	12.333	4.667

Kilde: TØI report 883/2007

Given the lack of standardisation and the many dimensions that characterise definitions of road accident black spots, the question must be asked: Are any of the definitions in any sense "better" than the others? To answer this question, one of course needs criteria specifying what constitutes a "good" definition of a road accident black spot. Such criteria have evolved over the past 20 years, following pioneering work by Hauer and Persaud (1984) who pointed out difficulties in reliably identifying black spots by using the recorded number of accidents as the only criterion. Overgaard Madsen (2005) discusses in great detail criteria for identifying black spots. He proposes that an adequate definition of a hazardous road location should satisfy three, or possibly four, criteria:

1. It should control for random fluctuations in the number of accidents.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

2. It should account for as many of the factors that are known to influence road safety as possible.
3. It should identify sites at which fatal and serious injury accidents are over-represented.
4. It should identify sites at which local risk factors related to road design and traffic control make a substantial contribution to accidents.

The first of these criteria suggests that the identification of a black spot should rely on the expected number of accidents, not the recorded number. In practice, this would appear to be difficult, since the expected number of accidents cannot be observed, but has to be estimated. However, a method has now been developed that permits the expected number of accidents to be estimated for a single location: the empirical Bayes method. By applying this method, it is in principle possible to identify hazardous road locations in terms of the number of accidents expected to occur in the long run at each such location.

The second and fourth criteria also suggest that the identification of road accident black spots should rely on the Empirical Bayes method, supported by a multivariate accident prediction model. By developing an accident prediction model, it is possible to account for a number of factors that explain systematic variation in the number of accidents, including traffic volume, various characteristics of road design and elements of traffic control (like speed limits). It is not realistic to expect an accident prediction model to include and accurately estimate the effects of all factors that influence the number of accidents; the factors that are not included in such models will typically be local risk factors, which, due to their site-specific nature, cannot be detected statistically in a multivariate model. These local factors may cause a site to have a higher expected number of accidents than predicted by an accident prediction model. The third criterion implies that the identification of black spots should either rely on fatal or serious accidents only, or assign a greater weight to these accidents than to slight injury accidents or property-damage-only accidents. This criterion is relevant to the extent that road safety policy seeks to prevent the most serious accidents.

It is useful to make a distinction between the definition of black spots and the identification of them. In some operational definitions of the concept, this distinction is blurred. However, as will become clear in the simulation study reported below, the distinction is important.

A state-of-the-art definition of a road accident black spot will be proposed in section 2.5. As far as is known, no European country at present consistently employs the state-of-the-art definition or state-of-the-art techniques for identifying black spots. It is therefore necessary to develop a step-by-step description of the state-of-the-art and support this description by studies that show that the proposed state-of-the-art approach performs better than alternative approaches. This description of the state-of-the-art is presented below in the form of a simulation and a review of some key studies.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

2.3 Limitations of traditional approaches: identifying black spots

The study presented here is strongly inspired by similar studies made many years ago by Hauer and Persaud (1984), Hauer and Quaye (1990) and Hauer, Quaye and Liu (1993). The study shows the limitations and pitfalls of identifying road accident black spots in terms of the recorded number of accidents only.

Assume that black spots are to be identified from a population of 1,000 sites. Table 5 lists this population, stratified into homogeneous groups with respect to the expected number of accidents. In practice, the expected number of accidents in a population of sites is a continuous variable, that cannot readily be stratified into homogeneous groups as shown in Table 5. The stratification is used for expository purposes only.

The first column shows the count of accidents. The distribution of sites by the number of accidents in each group was generated by assuming that accidents are Poisson distributed. This is equivalent to assuming that a perfect accident prediction model has been developed, which is able to explain all systematic variation in the number of accidents and discriminate perfectly between the groups formed in Table 5. In practice, of course, a perfect model is never developed; in the following it is assumed that only the column to the right in Table 5 is known. The groups are not known – they serve only to model the accident generating process if that process were perfectly known.

The column to the right shows the distribution of all 1,000 sites by the recorded number of accidents. The mean number of accidents for the population as a whole is 0.779. The variance is 2.003. If accidents were distributed at random, the variance would equal the mean (0.779). Thus, in this group 61% of the variation is systematic [(2.003 – 0.779)/2.003], 39% is random (0.779/2.003).

Assume that sites whose expected number of accidents is 4 are defined as black spots. As shown in Table 5, there were 50 black spots in the population. Accidents at the black spots represent 25% of the total number of accidents.

Table 5: A population of 1,000 sites stratified according to the expected number of accidents

	Groups according to the expected number of accidents					
Count	**0.2**	**0.5**	**1.0**	**3.0**	**4.0**	**Total**
0	532	61	37	5	1	636
1	106	30	37	15	4	193
2	11	8	18	22	7	66
3	1	1	6	22	10	40
4	0	0	2	17	10	29
5			0	10	8	18
6				5	5	10
7				2	3	5
8				2	1	3
9					1	1
Total	650	100	100	100	50	1000

Kilde: TØI report 883/2007

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

How can we best identify the black spots in this population? By assumption, the road safety manager knows only the rightmost column in Table 5, not the data for each of the groups. In other words: the recorded number of accidents and its variation between sites is known; the expected number of accidents for each site is unknown. Hence, to identify the black spots, the only option is to rely on the recorded number of accidents. It seems logical to identify sites that had 4 or more recorded accidents as black spots. By this criterion, 66 sites will be identified as black spots: 29 that have 4 accidents, 18 that have 5 accidents, and so on. Among these sites, however, only 28 will be true black spots, meaning their expected number of accidents is 4. These are the 28 sites that recorded at least 4 accidents whose expected number of accidents is also at least 4. These sites are found in the column for sites with an expected number of accidents equal to 4, and consist of sites that recorded 4 accidents (10), 5 accidents (8), and so on, up to 9 accidents. The remaining 38 sites will be false positives. These sites consist of sites that recorded 4 or more accidents, but have an expected number of accidents less than 4. We may now define four categories of sites:

1. Correct positives: These are sites at which the expected number of accidents exceeds the critical value selected and the recorded number of accidents exceeds the same critical value.
2. False positives: These are sites at which the expected number of accidents does not exceed the critical value selected, but the recorded number of accidents does exceed this value as a result of random variation.
3. Correct negatives: These are sites at which both the expected and recorded number of accidents are lower than the critical value selected.
4. False negatives: These are sites at which the expected number of accidents exceeds the critical value selected, but the recorded number of accidents does not, due to random variation.

Table 6 shows the number of sites in these groups as a function of the recorded number of accidents used to identify black spots.

Table 6: Number of correct negatives, false negatives, correct positives and false positives as a function of the critical number of accidents. Sites with an expected number of accidents of 4 are defined as black spots

Critical number	Correct negatives	False negatives	Correct positives	False positives	Total sites identified
1	635	1	49	315	364
2	823	5	45	127	172
3	883	12	38	67	105
4	912	22	28	38	66
5	931	32	18	19	37
6	941	40	10	9	19
7	946	45	5	4	9
8	948	48	2	2	4
9	950	49	1	0	1

Kilde: TØI report 883/2007

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

If all sites that have recorded an accident at all are included, 364 sites will be identified. The great majority of these, 315 sites, will be false positives. There will be 49 correct positives and 1 false negative. Thus even if the criterion is set as low as it could possibly be – a single accident – there will still be one black spot that is not detected.

If more stringent critical values are adopted, fewer sites will satisfy them. A growing proportion of the sites identified will be correct positives, but this is accomplished at the cost of a growing number of false negatives. At nine accidents, only a single site is identified, but 49 sites go undetected.

It can be seen that no criterion for identifying hazardous road locations is perfect. The reason is very simple. We cannot observe the expected number of accidents. We can only observe the recorded number of accidents, which is always partly the outcome of chance, partly the outcome of very many factors that systematically influence the expected number of accidents.

The performance of the various criterion values can be assessed quantitatively in terms of screening performance criteria developed in epidemiology (Deeks 2001, Rothman and Greenland 1998). Two of the most common criteria for diagnostic tests are sensitivity and specificity. They are defined as follows:

$$\text{Sensitivity} = \frac{\text{Number of correct positives}}{\text{Total number of positives}}$$

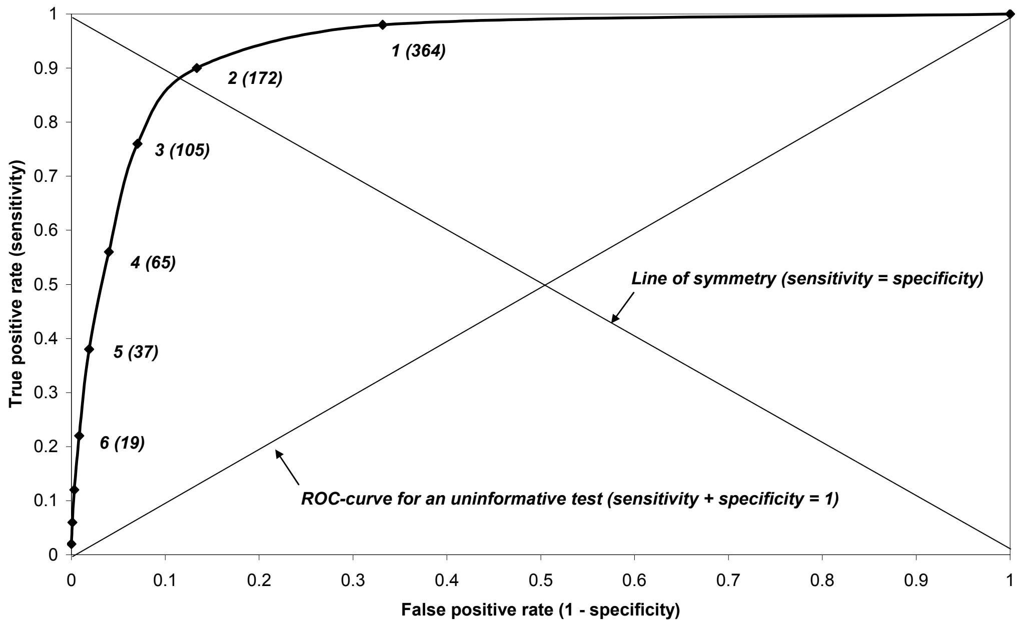
$$\text{Specificity} = \frac{\text{Number of correct negatives}}{\text{Total number of negatives}}$$

The total number of positives equals the number of correct (true) positives plus the number of false negatives. With reference to Table 6, the sensitivity of using 4 accidents as the diagnostic criterion is 28/50 = 0.56. The specificity of this criterion is 912/950 = 0.96. The performance of different values for the critical number of accidents used to identify a black spot can now be assessed in terms of a receiver operating characteristic curve (ROC-curve). Such a curve, derived from the data in Table 4, is shown in Figure 4.

The false positive rate is plotted along the abscissa. This is equal to 1 minus specificity. The true positive rate (sensitivity) is plotted on the ordinate. If the diagnostic test discriminates well, the ROC-curve will rise steeply, close to the ordinate and flatten out near the top of the diagram. If the diagnostic test is uninformative, the ROC-curve will follow the diagonal line indicated in Figure 4.

It is desirable to minimise the false positive rate and to maximise the true positive rate. This involves a trade-off; one may diminish the false positive rate by accepting a lower true positive rate, and vice versa. The optimal criterion is the one that maximises the sum of sensitivity and specificity. For Figure 4, this is to treat all sites with 2 or more accidents as potential black spots. This is marginally better than using 3 accidents as the criterion.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961



Kilde: TØI report 883/2007

Figure 4: ROC-curve for detecting road accident black spots. Simulated data

In practice, the criterion defining a black spot is rarely, if ever, based on an evaluation of the diagnostic performance of the criterion. Ideally speaking, an optimal criterion of deviance ought to be chosen. On the other hand, practical considerations may prevent this. In the example above, if all sites with 2 or more accidents are treated as black spots, 172 sites would be identified, of which 127 would be false positives. This would create a considerable amount of work in performing accident analysis for the purpose of diagnosing problems at each site and, again ideally speaking, reliably identify the true and false black spots. Thus, the choice of a criterion for identifying black spots cannot be based on a statistical criterion only. No statistical criterion can reliably identify only correct black spots, and include all of them, as the criterion would always be applied to a population of sites containing a mixture of random and systematic variation in the number of accidents. Indeed, the very idea of selecting black spots from a population of sites containing random variation only is a contradiction in terms, as a true black spot should always be defined as having a higher expected number of accidents than other, similar sites. In a completely homogeneous population, by definition no such sites would exist. One must therefore always identify black spots in a heterogeneous population, that contains unexplained systematic variation in accident counts, with random variation on top of this.

This means that any criterion will be imperfect: Sites identified as black spots will always contain a mixture of correct positives and false positives. Besides, there will always be a number of false negatives. The number of sites that are true or false black spots will almost never be known.

These difficulties are compounded if black spots are identified by applying a sliding window method, that determines the location of a black spot according to the location of the accidents, rather than by sampling locations from a known

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

sampling frame allowing sites to be enumerated. This was shown empirically by Elvik (1988). Table 7 reproduces some of his findings.

The first column shows the distribution of 100 1-kilometre road sections by the number of accidents. The mean number of accidents was 2.44. The variance was 4.37. The distribution does not differ significantly from a negative binomial distribution. The next column shows the results of applying a sliding window of length 1-kilometre to the same 100 kilometres of road as shown in the first column. It is seen that the tail of the distribution becomes substantially longer. The number of road sections with 0 accidents now consists of a mixture of "rump" sections of varying length interspersed among the sections where accidents have been recorded. Some of these rump sections will be shorter than 1 kilometre, some longer. If we take sections with 5 or more accidents as black, there are 15 such sections in the population of fixed sections, but 19 sections according to the sliding window approach. Using the sliding window artificially inflates the number of black sections, and makes each section look more black (i.e. having a higher recorded number of accidents) than it really is. The findings are the same when fixed and sliding 4-kilometre sections are compared. The number having 5 or more accidents is 14 when fixed sections are used, 19 when sliding sections are used.

Table 7: Effects of using a sliding window approach on the number of black spots identified

Number of accidents	**Fixed 1-kilometre sections, 4 years**	**Sliding 1 kilometre sections, 4 years**	**Fixed 4-kilometre sections, 1 year**	**Sliding 4-kilometre sections, 1 year**
0	19	Not defined	11	Not defined
1	19	9	25	9
2	19	13	23	12
3	18	11	16	16
4	10	11	11	12
5	7	7	6	8
6	3	4	5	5
7	3	3	3	5
8	1	0		0
9	0	2		0
10	0	2		1
11	1	0		
12		0		
13		0		
14		1		
Total sections	100	65	100	68
Mean per section	2.44	3.87	2.44	3.59
Variance	4.37	6.59	3.19	3.57

Kilde: TØI report 883/2007

Theoretical work by Hauer and Quaye (1990) confirms these findings. Hauer and Quaye applied a sliding window approach to a fictitious population of sites of which 900 sites had an expected number of accidents of 1, 90 had an expected number of accidents of 2 and 10 had an expected number of accidents of 3. Within

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

each group, the recorded number of accidents was assumed to be Poisson distributed around the mean value. A critical value of 5 accidents was used to identify deviant sites.

If fixed sections are used, it can be estimated that 10 sites will be identified as positives. If, however, a sliding window is used, 34 sites will be identified, of which the great majority will be safer than average. Thus, using a sliding window greatly inflates the number of false positives identified as black spots.

Applying a sliding window of standard length 1 (arbitrary units) to the fictitious data in Table 5 confirms that this method for identifying black spots artificially inflates their number. Using 4 accidents inside the window as critical number, 113 black spots were identified. 42 of these were correct positives, the rest were false positives (71 sites in total). When fixed road sections are used, a critical value of 4 accidents identifies 66 sites, of which 28 correct positives and 38 false positives. Thus, use of a sliding window has the advantage of identifying more correct positives (42 versus 28), but the disadvantage of identifying more false positives (71 versus 38). In terms of the epidemiological criteria, the sliding window approach (critical value 4) has a specificity of 0.925 and a sensitivity of 0.840. The corresponding values for fixed sections (critical value 4) are 0.96 and 0.56, respectively.

2.4 Limitations of traditional approaches: analysis of black spots

The limitations of traditional approaches to the analysis of black spots are reviewed by Elvik (2006). The following section draws on his review.

No matter how accident black spots are identified statistically, the identification will not be perfect in the sense that all sites identified have a high expected number of accidents and none of them have a high recorded number of accidents due to chance mainly. While progress has been made in developing techniques that keep the number of false positives down, any set of hazardous road locations identified statistically will contain both true and false positives. Ideally speaking, a detailed analysis of accidents at sites that are identified as hazardous should be able to discriminate between the sites that are true positives and those that are false positives. However, research sheds doubt on this assumption.

Danielsson (1988) shows that one commonly used criterion for identifying a truly hazardous road location, namely the overrepresentation of a particular type of accident is vulnerable to regression-to-the-mean bias, because overrepresentation could be attributable mainly to chance. Suppose, for example that at site 1, there were 1 accident of type A, 2 accidents of type B, 1 accident of type C, and 1 accident of type D. Since no type of accident is clearly dominant, it might be concluded that site 1 is not a true positive. At site 2, on the other hand, there were 4 accidents of type A, 1 accident of type B and no accidents of types C or D. Since accidents of type A seem to be dominant at site 2, it is tempting to conclude that site 2 is a true positive. If a treatment designed to reduce accidents of type A is then introduced, it will often be observed that accidents of that type are reduced more than accidents of other types. In principle, such an apparently systematic trend could be largely attributable to chance.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

In a similar vein, Jarrett, Abbess and Wright (1988) compared regression-to-the-mean for a sample of high-accident sites in London that had undergone analysis and been selected for treatment, but where the treatment had not been implemented, to a sample of sites that had not been selected for treatment. If analysis and selection for treatment successfully identifies true black spots, one would expect the regression-to-the-mean effect to be smaller at these sites than at similar sites that had not been analysed and selected for treatment. The study did, however, not find such a difference. The size of regression-to-the-mean was very similar in the two samples.

It can be shown by means of simulation that regression-to-the-mean will be greater at false black spots than at true black spots. For this purpose, the expected regression-to-the-mean effect has been estimated for true and false black spots based on the hypothetical data in Table 5. Results are shown in Table 8.

Suppose that a count of 4 accidents is used to identify black spots. As shown in Table 5, 66 locations will then be identified. 28 of these locations will be correct positives, i.e. their expected number of accidents equals 4. The remaining 38 locations will be false positives, i.e. their expected number of accidents is less than 4.

Table 8: Expected regression-to-the-mean for true and false black spots based on simulated data

	False positives (mean < 4)			Correct positives (mean = 4)		
Critical accident count	Mean number of accidents	Mean expected number of accidents	Regression-to-the-mean (%)	Mean number of accidents	Mean expected number of accidents	Regression-to-the-mean (%)
3	4.06	2.67	34	4.68	4.00	15
4	4.89	2.86	41	5.29	4.00	24
5	5.79	3.00	48	6.00	4.00	33
6	6.67	3.00	55	6.80	4.00	41
7	7.50	3.00	60	7.60	4.00	47
8	8.00	3.00	62	8.50	4.00	53
9				9.00	4.00	56

Kilde: TØI report 883/2007

The mean recorded number of accidents for the 28 correct positives will be 5.29. Since the assumption has been made that accidents are randomly distributed around the mean value, the number of accidents at these sites can be expected to regress to the mean value of 4. The expected regression-to-the-mean effect is 24 percent. For the false positives, the mean recorded number of accidents will be 4.89. This is expected to regress to a mean of 2.86. The expected regression-to-the-mean effect in this case is 41 percent.

Thus, the finding of Jarrett, Abbess and Wright casts serious doubt on whether accident analysis and selection for treatment was able to successfully discriminate between the false and correct positives.

A commonly applied criterion to discriminate between true and false black spots is the presence of a dominant accident pattern. A dominant accident pattern is characterised by the overrepresentation of a particular type of accident. It is therefore of some interest to probe whether there is any difference in the

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

regression-to-the-mean effect between hazardous road locations that have a dominant accident pattern and those that do not.

In order to test this, data for 1-kilometre road sections on national roads in Norway were used. Data for two 4-year periods were used. Sections that during the first period recorded 4 or more accidents were identified as hazardous. A distinction was made between the following types of accident:

- Head-on collisions
- Single vehicle running off the road
- Rear-end collisions
- Intersection accidents
- Pedestrian accidents

For the road system as a whole, single vehicle running off the road made up 36% of the accidents, rear-end collisions 25%, head on collisions 10%, intersection accidents 10% and pedestrian accidents 5%. Various other types of accident, not specified here, made up the rest.

A hazardous road section was considered to have a dominant accident pattern if any the types of accidents listed above represented at least 60% of all accidents on the section. Sections where none of the accident types represented as much as 60% of all accidents were not considered to have a dominant accident pattern. Table 9 shows the mean recorded number of accidents for sites with and without a dominant accident pattern during the two 4-year periods used in analysis.

A total of 700 sections were identified as hazardous, i.e. they had at least 4 injury accidents in the first 4-year period. 292 of these sections had a dominant accident pattern. i.e. one type of accident accounted for at least 60% of all accidents. 408 sections did not have a dominant accident pattern.

The difference between these groups with respect to regression-to-the-mean during the next four years was small, 22% for the group with a dominant accident pattern compared to 26% for the group without a dominant accident pattern. This suggests that the dominance of a particular type of accident does not necessarily persist over time. The pattern is broadly speaking the same for all types of accident. In most cases, running-off-the-road accidents forming an exception, there is a tendency for the regression-to-the-mean effect to be slightly smaller when there is dominant accident pattern than when this is not the case.

The probability that an accident pattern characterised by the dominance of a certain type of accident may arise by chance is not negligible. By applying a binomial model (each accident is either of the selected type or of any other type), it can be estimated that 138 of the 292 sites with a dominant accident pattern can be expected to arise from chance alone. Hence, a sizable proportion of these sites are false positives and a large regression-to-the-mean effect should be expected.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 9: Regression-to-the-mean at hazardous road sections in Norway for sections with and without a dominant accident pattern

		Mean number of injury accidents per section		
Group of sections	**Number of sections**	**First four years**	**Second four years**	**Regression-to-the-mean (%)**
All sections	700	6.82	5.16	-24%
Sections with dominance of any type of accident	292	6.94	5.39	-22%
Sections without dominance of any type of accident	408	6.74	5.00	-26%
Dominance of rear-end collisions	172	8.13	6.84	-16%
No dominance of rear-end collisions	528	6.39	4.62	-28%
Dominance of running-off-the road	76	4.96	2.47	-50%
No dominance of running-off-the road	624	7.05	5.49	-22%
Dominance of intersection accidents	33	5.73	4.97	-13%
No dominance of intersection accidents	667	6.87	5.17	-25%

Kilde: TØI report 883/2007

Harwood et al (2002) point out that some sites with a high number of accidents do not have readily identifiable accident patterns. A given deficiency in highway design or traffic control can contribute to accidents at one site, while at another site with a similar deficiency, there are no accidents or no clear pattern of accidents associated with the deficiency. Finally, a given deficiency can contribute to different accident types. This suggests that an analysis of accidents designed to identify true black spots must go beyond merely identifying a dominant accident pattern.

A new approach to accident analysis, designed to be better able to discriminate between false and true black spots will be discussed in section 2.7.

Elvik (1997) compared the findings of studies that have evaluated the effects of black spot treatment, depending on how well studies controlled for various confounding factors. He found that the effect attributed to treatment declined as studies controlled for more confounding factors. In the best controlled studies (controlling for regression-to-the-mean, long-term trends and accident migration), the effect on safety attributed to treatment was zero. This suggests that the treatments did not successfully address risk factors contributing to accidents, or that the accidents were mainly the result of chance variation.

Brüde and Larsson (1982) point out the crucial role that analyst expectations may play in the analysis of accidents at black spots: “One might hope that the accident analysis together with on-site inspection should enable us to decide whether a high recorded number of accidents is due to chance or to deficiencies in design or traffic control. In practice this may not be possible. It is nearly always possible to point out some deficiency – in particular this is so when we know that a large number of accidents have occurred.” Thus, even if no clear pattern is found, or no specific risk factors contributing to accidents can be identified, the temptation may be almost irresistible to conclude that there has got to be something wrong about the place, since there have been so many accidents there. It is almost always

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

possible to point out some deficiency in design or traffic control. It is well known from psychological research that analyst or experimenter expectancies regarding the outcome of a study can exert a major influence on the conclusions drawn from it (Rosenthal and Rubin 1978). In summary, research points to the possibility that:

1. The frequently used criterion for a true black spot – the presence of a clearly identifiable pattern of accidents characterised by the dominance of a particular type of accident – may not effectively separate true from false black spots.
2. Sites that have been analysed and selected for treatment do not necessarily have a higher long-term expected number of accidents than similar sites that have not been analysed and selected for treatment.
3. Treatments do not appear to always successfully address risk factors contributing to accidents.
4. Analysts may find it difficult to resist the temptation to conclude that a site at which a high number of accidents have been recorded must have some deficiency in design or traffic control, even if no clear accident pattern pointing to such deficiencies can be identified. The idea that an analysis might be inconclusive is unappealing and resisted.

In short, an approach to accident analysis is needed that provides clearer criteria for identifying true black spots, recognises the possibility that analysis might be inconclusive, and minimises the role of analyst expectancies.

2.5 Theoretical definition of a black spot

Most currently employed definitions of road accident black spots are theoretically unsatisfactory, mainly because they are not clear about whether black spots should be defined in terms of the expected number of accidents or in terms of the recorded number of accidents. From a theoretical point of view, there can be no doubt at this point: We are seeking to identify sites whose expected number of accidents is abnormally high, not sites that happen to have recorded a high number of accidents due to chance. Thus, the following theoretical definition of an accident black spot is proposed.

A road accident black spot is any location that:

1. Has a higher expected number of accidents,
2. Than other similar locations,
3. As a result of local risk factors.

Black spots are identified from a known population of sites. Similarity is assessed in terms of the values for explanatory variables in accident prediction models. Thus, two sites can be treated as similar if, for example, they have the same traffic volume, the same speed limit, the same number of lanes, the same number of junctions per kilometre, and so on. When identifying black spots, comparison to other similar locations ensures that the explanatory factors included in accident prediction models are controlled for. Controlling for these factors is essential, as we do not want to identify a site as a black spot simply because it has a higher

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

traffic volume than another site. It is normal for the number of accidents to increase as traffic volume increases.

The theoretical definition given above implies that the only method that can reliably identify black spots is one that facilitates the identification of the contribution of the three main factors to the expected number of accidents for a single site. The Empirical Bayes method is one such method, but it is possible that other methods could provide adequate approximations. EB-estimates of safety do, however, allow the relative contributions of random variation, general factors (included in the accident prediction model) and local factors to the observed number of accidents to be identified.

2.6 Statistical identification of black spots: a comparative analysis

If one accepts the theoretical definition of a road accident black spot proposed above, a number of expectations regarding empirical methods for identifying black spots follow. These include:

1. The EB-method is likely to give the most reliable identification of black spots, i.e. minimise the number of false positives and false negatives.
2. Black spots are likely to be most reliably identified in terms of the relative contribution of local risk factors to the recorded number of accidents.

To evaluate whether these implications hold in practice, a comparative study was made of the following criteria for statistical identification of black spots:

1. ***Upper tail accident count***: Sites whose recorded number of accidents belonged to the upper 2.5% of the distribution during the first four years were identified as hazardous road locations. Sites that continued to belong to the upper 2.5% in the second period were classified as correct positives. Sites that dropped out of the list were classified as false positives, new sites entering the list were classified as false negatives. The procedure was repeated using the upper 1% and the upper 5% of the distribution as criteria.
2. ***A critical accident rate***: A period of four years was used. Accident rate was defined as the number of injury accidents per million vehicle kilometres. Sites that had the 2.5% highest values for the accident rate were classified as black spots (irrespective of the number of accidents at these sites). The accident rate for the same sites were observed for the next four years. Sites that continued to belong to the top 2.5% were classified as correct positives. Sites whose accident rate dropped below the top 2.5% were classified as false positives. Sites that did not belong to the top 2.5% in the first four years, but did so in the second four years were classified as false negatives. The procedure was repeated using the upper 1% and upper 5% of the distribution as criteria.
3. ***A critical rate and number of accidents***: A period of four years was again used. Sites that recorded a number of accidents greater than the upper 2.5%, 1% or 5% values in the population of sites, and had a higher than

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

average accident rate were classified as black spots. Average accident rate referred to the overall average for the whole population of sites, not the average for the upper 2.5%, 1% or 5%. Sites that during the second four year period continued to satisfy both criteria were classified as correct positives. Sites that in the second four year period failed to satisfy one or both criteria were classified as false positives. Sites that did not satisfy the criteria in the first four years, but did so in the second four years, were classified as false negatives.

4. ***Upper 2.5% EB-criterion***: For each site, the EB-estimate of the expected number of accidents for that site was developed, based on four years of data (and an accident prediction model). Sites with the 2.5% highest estimates were classified as black spots. If the EB-estimate for these sites in the second four year period remained in the upper 2.5%, the sites were classified as correct positives. Sites that dropped out of the upper 2.5% were classified as false positives, new sites that entered the list were classified as false negatives. The procedure was repeated using the upper 1% and 5% as critical values.
5. ***The EB-dispersion criterion (potential accident reduction)***: For each site, an EB-estimate of safety was developed (based on an accident prediction model). The recorded number of accidents, for sites that had a higher recorded number of accidents than the number predicted according to the model, was decomposed into contributions from three factors: (a) Randomness, (b) General risk factors (included in model), and (c) Local risk factors. Sites were sorted by the contribution from local factors; sites at the top 2.5% (for the whole population of sites) are classified as black spots, provided the recorded number of accidents was 4 or more. Sites that remained in the upper 2.5% in the second period were classified as correct positives. Sites that dropped out were treated as false positives, new sites entering were treated as false negatives. The procedure was repeated using 1% or 5% as the critical values, but keeping the critical number of accidents constant

Data for 1-kilometre sections of national roads in Norway were used to assess all definitions. Data for two periods of four years were used. These data covered the period from 1997 to 2004. Hazardous road locations were identified on the basis of data referring to the first four years (1997-2000). To assess whether the hazardous road locations were true or false positives, data referring to the second four year period were used (2001-2004). The idea was that true positives will persist in having a bad safety record, whereas false positives will regress towards a more normal safety record. There will also be some false negatives, i.e. sites not detected in first four years that are detected in the second four years.

An accident prediction model was developed based on data referring to the first four years. The model included the following explanatory variables: AADT, speed limit (km/h), number of lanes, number of intersections per kilometre of road and a dummy for trunk roads. The model was of the form:

$$E(\lambda) = \alpha Q^{\beta} e^{\sum \gamma_i x_i} \tag{7}$$

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

The estimated normal number of accidents, E(λ), is a function of traffic volume, Q, and the other explanatory variables, X_i (i = 1, 2, 3, … n). The data referred to road sections of 1 kilometre. Location reference data were available for all sections. The number of sections included in the study was 19,623. Sections shorter than 1 kilometre and sections that did not exist in the whole period were not included. Table 10 presents the results of the analysis.

Table 10: Comparison of alternative definitions of road accident black spots in terms of epidemiological criteria.

Identification criterion	Correct negatives	Correct positives	False negatives	False positives	Sensitivity	Specificity
Top 1 % of distribution						
Accident count	19272	134	109	108	0.551	0.994
Accident rate	19232	16	188	187	0.078	0.990
Accident rate and count	19340	86	94	103	0.478	0.995
EB-estimate of accidents	19378	130	53	62	0.710	0.997
EB dispersion criterion	19311	62	121	129	0.339	0.993
Top 2.5 % of distribution						
Accident count	18788	285	262	288	0.521	0.985
Accident rate	18726	53	418	426	0.113	0.978
Accident rate and count	18928	186	236	273	0.441	0.986
EB-estimate of accidents	18981	338	152	152	0.690	0.992
EB dispersion criterion	19070	105	195	253	0.350	0.987
Top 5 % of distribution						
Accident count	18065	464	526	568	0.469	0.970
Accident rate	17838	144	805	836	0.152	0.955
Accident rate and count	18308	307	474	534	0.393	0.972
EB-estimate of accidents	18429	692	235	267	0.746	0.986
EB dispersion criterion	18989	136	219	279	0.383	0.986

Source: TØI-report 883/1007

The total number of positives equals the number of correct (true) positives plus the number of false negatives. The total number of negatives equals the number of correct negatives plus the number of false positives.

To compare the performance of the different techniques for identifying hazardous road sections, the values of sensitivity and specificity can be added, since a good diagnostic test should score high on both criteria. Thus, for the upper 1 % of the distribution, using the accident count as criterion, gives a total diagnostic score of 0.551 + 0.994 = 1.545. For the EB-estimate of the expected number of accidents, the corresponding value is 1.707.

The empirical Bayes technique is found to perform best at all levels of stringency (1 %, 2.5 %, 5 %). Using the accident count performs second best. It is not surprising that the EB-criterion, which is the only one that is strictly based on estimates of the expected number of accidents for each site, performs better than the other criteria for identifying hazardous road locations. Somewhat surprisingly,

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

the EB dispersion criterion, which has strong theoretical arguments in its favour, does not perform well. The major problem associated with this criterion appears to be a large number of false positives. This may be the result of a temporal instability in the contribution of local risk factors to accidents. Perhaps the contribution of these factors shifts over time, making an identification of hazardous road locations in terms of these factors less reliable than an identification which is not based on the contribution of a specific source to the variation in the expected number of accidents.

An identical analysis was made using data for 3407 sections on Portuguese roads. Most of these sections were 250 metres long. The results were somewhat more mixed. The EB-criterion performed best for the upper 5% tail of the distribution. For the upper 2.5% of the distribution, the accident count criterion and the accident rate and number criterion tied as best performers. This suggests that a high recorded number of accidents in Portugal is associated with a high and stable accident rate, which is not always the case in Norway. For the upper 1% of the distribution, the accident rate was the best diagnostic criterion in Portugal – again suggesting that a high accident rate implies a high expected number of accidents.

It is thus not always the case that the empirical Bayes criterion most accurately identifies hazardous road locations. It did, however, do so in the majority of cases used in this test, as well as in earlier research reported by Persaud et. al. (1999). There are strong arguments for relying on EB-estimates even if their diagnostic performance is not necessarily superior in every case to other methods that can be used to identify hazardous road locations statistically. The main arguments favouring the EB-approach are:

1. EB-estimates permits the contributions to the expected number of accidents attributable to general factors, local risk factors and randomness to be identified statistically.
2. By virtue of this decomposition of contributing factors, EB-estimates predict how large the regression-to-the-mean effect will be. If these predictions tend to be correct, accident analyses should allow for them.
3. EB-estimates have known standard errors. This fact can be exploited in the profiles-and-peaks method used to identify longer sections of road for detailed engineering study (see further discussion in Chapter 4).

Points 2 and 3 above are discussed more in detail in Chapter 4.

2.7 A new approach to the analysis of black spots

The approach to accident analysis for hazardous road locations proposed here keeps all the elements of current approaches, but adds new elements (Elvik 2006). It is recognised that this will make the analysis more demanding, but it will hopefully also make it more conclusive.

The first stage of analysis is identical to the current practice of searching for patterns in accident data. It is proposed to formalise this search by relying on statistical tests and pattern recognition methods, as indicated by Kononov (2002).

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Consider the data presented in Table 11 as an example. These data are fictitious, in order to make some of the points of the approach clearer.

Eight accidents were recorded at a hazardous road location. Analysis shows that five of the eight accidents were pedestrian accidents, whereas one would normally expect only one in eight accidents to involve pedestrians. If the distribution of accidents by type is modelled as a binomial trial (each accident is either of the specified type or any other type), it is found that recording five pedestrian accidents in a total of eight accidents is a highly unlikely outcome. The normal probability of a pedestrian accident is 0.125 (the probability of a non-pedestrian accident is 0.875). The probability of observing 5 pedestrian accidents out of 8 is only 0.0011, given that one would expect to observe 1 pedestrian accident out of 8. Similar tests are reported at the bottom of Table 11 for each variable recorded. For each variable, the probability of observing the overrepresented value of that variable is estimated on the basis of the outcome one would normally expect to find.

On the whole, the predominant accident pattern found in table 11, pedestrian accidents occurring at night on a wet road surface, suggests that local risk factors related to the amount of pedestrian traffic, road surface friction and visual obstructions may be present at the location. In table 11, the number of accidents normally expected to occur according to all logically possible combinations of values for road user group (pedestrian ($p = 0.125$) or other ($p = 0.875$)), road surface condition (wet ($p = 0.25$) or dry ($p = 0.75$)) and presence of alcohol (yes = 0.125, no = 0.875) has been estimated and is compared to the actual distribution of accidents. It is seen that the combination pedestrian, wet road and alcohol involved occurs much more frequently than one would expect in a random sample of 8 accidents (confer Table 12).

Despite this, a more careful investigation would be needed in order to determine whether the factors suggested are actually responsible for the abnormally high number of pedestrian accidents at this particular location. Accident analysis at hazardous road locations amounts to proposing hypotheses based on known data, which means that the data that generated the hypotheses cannot also be used for testing them. Thus, the principal results of an analysis of accidents should be regarded as hypotheses only, to be tested in subsequent steps of the analysis. These steps can be outlined as follows:

1. For each hazardous road location, find a safer-than-average comparison location, matched as closely to the hazardous road location as possible with respect to variables included in an accident prediction model used to predict the normal number of accidents.
2. For each matched pair of sites, search for local risk factors or safety factors from a list of factors drawn up on the basis of the analysis of accidents at the hazardous road location.
3. Blind analysts to accident records. Analysts should not know which site was hazardous and which site was safer than average.

The use of this approach is shown in Table 13. Hazardous and safe sites are matched in pairs according to the values observed for the variables included in an accident prediction model. Two matched pairs are shown in table 13. Once the

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

pairs have been formed, each site is inspected and data collected regarding local risk factors. A sample of such data, not necessarily exhaustive, is shown in table 12.

In case of the first pair of sites, it was found that wet road surface friction was significantly worse, that there were more pedestrians crossing the road, and more sources of visual obstruction at the hazardous site than at the safe site. This information confirms the hypotheses regarding contributing factors proposed on the basis of the analysis of accidents. The analysis has therefore successfully identified local risk factors. Keep in mind that the analysts identifying risk factors should be blinded to accident records, to prevent their knowledge of accident records from biasing their observations.

The other case shown in table 13 was less successful. It turned out that there were no differences between the hazardous and the safe site with respect to the risk factors surveyed. Hence, accidents must be attributed to other risk factors, for example a widespread violation of speed limits or other traffic control devices, or to chance fluctuations.

Based on this logic, criteria can now be proposed regarding the conclusion to be drawn from an analysis of accidents and risk factors at hazardous road locations and matched comparison sites. These criteria are shown in Table 14. A distinction can be made between four cases. In the first case, factors associated with accidents at the hazardous road location are identified in the accident analysis, and the hypotheses regarding contributing risk factors are supported, meaning that these risk factors are found to be more clearly present (assume less favourable values) at the hazardous road location that at the matched comparison site. In this case, it is reasonable to conclude that the hazardous road location is a true positive, i.e. a site that has a higher expected number of accidents than similar sites, due to local risk factors.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 11: Hypothetical results of accident analysis at a hazardous road location

Accident number	Type of accident	Time of day	Road surface	Vehicles involved	Alcohol involved	Excessive speed	Failure to see
1	Pedestrian	11 PM	Wet	Car	Yes, pedestrian	Yes	Yes
2	Rear-end	10 AM	Wet	Truck	No	No	No
3	Rear-end	5 PM	Dry	Car	No	No	No
4	Pedestrian	8 PM	Dry	Car	No	Yes	No
5	Pedestrian	9 PM	Wet	Car	Yes, pedestrian	No	Yes
6	Pedestrian	11 AM	Wet	Car	No	Yes	Yes
7	Overturning	1 PM	Dry	Motorcycle	No	Yes	No
8	Pedestrian	11 PM	Wet	Truck	Yes, pedestrian	No	Yes
Key finding	5 pedestrian	4 in evening	5 on wet road	2 trucks	3 with alcohol	4 speeders	4 did not see
Normal value	1 pedestrian	2 in evening	2 on wet road	1 truck	1 with alcohol	3 speeders	2 did not see
P-value (binomial)	0.0011	0.0865	0.0231	0.1963	0.0561	0.2112	0.0865
Predominant accident pattern	The predominant type of accident is a pedestrian accident at night on a wet road surface, in which the parties did not see each other. Some over-involvement of alcohol among pedestrians.						

Table 12: Comparison of the number of accidents normally expected to occur and the actual number of accidents according to three characteristics associated with accidents

Road user involved	Road surface condition	Alcohol involved	Expected number	Observed number	Ratio observed/expected
Pedestrian	Wet	Yes	0.03	3	96.0
		No	0.22	1	4.6
	Dry	Yes	0.09	0	0.0
		No	0.66	1	1.5
Other	Wet	Yes	0.22	0	0.0
		No	1.53	1	0.7
	Dry	Yes	0.66	0	0.0
		No	4.59	2	0.4
Total	Total	Total	8	8	1.0

Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

In the second case, the accident analysis identified a clear pattern, suggesting that specific risk factors contributed to the accidents, but the subsequent matched-pair analysis of these risk factors does not support these hypotheses. In this case, the analysis is inconclusive. The accidents may be more closely associated with other risk factors than those examined, but they may also be the result of random fluctuations mainly. In a case like this, it is tempting to carry on the analysis by examining one risk factor after the other, until one or more factors are found to be associated with accidents. This practice should be discouraged. It amounts to data mining, which, if carried out long enough, will always turn up something that looks systematic. Surely, among the hundreds of risk factors that contribute to accidents, one of them may by chance seem to be associated with the accidents recorded at a particular location. It is precisely to guard against this sort of data mining that the examination of risk factors should be limited to a specific set of risk factors that have been identified in the accident analysis as potentially contributing to the specific pattern of accidents observed at a site.

In the third case, the first stage of analysis is "unsuccessful", in the sense that no clear pattern is found and no hypotheses regarding specific risk factors contributing to the accidents can be proposed. The site could, as pointed out by Harwood et al (2002), nevertheless be a true black spot. However, for it to be so, the accidents would have to be mainly associated with fairly general risk factors, i.e. risk factors that are more or less associated with all accidents, and that do not necessarily result in the predominance of a particular type of accident. Risk factors that may contribute to any type of accident include speed, road surface friction, lateral placement of vehicles and the following distances of vehicles. These risk factors are always present, but they could form an unfortunate combination at a particular location. To test if this is the case, one could compare observed values for the general risk factors at a hazardous road location to a matched comparison site. If the values observed were, in general, less favourable at the hazardous road location than at the comparison site (higher speed, less friction, lateral placement giving a smaller safety margin to either the edge of the road or the median, shorter following distances), it would seem reasonable to conclude that the site is a true black spot.

The fourth case is identical to the third, except that no evidence is found indicating that the general risk factors are contributing to the accidents at the hazardous road location. In this case, it is reasonable to conclude that the site is a false positive, since there is no discernible pattern in accidents, and since no risk factors can be found to be associated with the accidents.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 13: Verification of traditional accident analysis by identification of risk factors contributing to accidents

	Case 1: Local risk factors successfully identified		Case 2: Local risk factors not identified	
	Hazardous	**Safe**	**Hazardous**	**Safe**
Matching variables	AADT, speed limit, number of lanes, number of intersections, trunk road status = nearly same values observed for both sites		AADT, speed limit, number of lanes, number of intersections, trunk road status = nearly same values observed for both sites	
Local risk factors (sample):				
Road surface friction – dry	0.70	0.82	0.78	0.77
Road surface friction – wet	0.25	0.48	0.47	0.49
Pedestrians crossing per day	2,500	1,000	1,200	1,250
Sources of visual obstruction	5	2	3	3
Minimum sight distance (m)	100	155	110	115
Driveways per km of road	2	2	0	0
Public bar nearby	Yes	No	No	No
Accident records:	8 accidents in total, of which 5 involving pedestrians on a wet road surface	0 accidents	7 accidents in total, but no clear pattern	0 accidents

Kilde: TØI report 883/2007

Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 14: Conclusion drawn from analysis of accidents and risk factors at hazardous road locations and matched comparison locations

Results of accident analysis	Results of matched comparison of risk factors	Conclusion from analysis
A pattern of characteristics associated with accidents is found – hypotheses regarding contributing risk factors are proposed	The hypothesised risk factors are found to be more clearly present at the hazardous location than at the matched comparison location	The hazardous road location is ***likely to be a true positive site***, with a high long-term expected number of accidents
A pattern of characteristics associated with accidents is found – hypotheses regarding contributing risk factors are proposed	The hypothesised risk factors are ***not*** found to be more clearly present at the hazardous road location than at the matched comparison location	The analysis is ***inconclusive***; accidents are not found to be the result of the risk factors examined and could therefore be the result of chance or of risk factors not examined
No clear pattern of characteristics associated with accidents is found – hypotheses regarding contributing risk factors are difficult to develop	In a matched comparison with respect to a few risk factors associated with all accidents (speed, friction, lateral placement, following distance) these factors are found to have less favourable values for the hazardous site than the matched comparison site	The hazardous road location is ***likely to be a true positive site***, with a high long-term expected number of accidents
No clear pattern of characteristics associated with accidents is found – hypotheses regarding contributing risk factors are difficult to develop	In a matched comparison with respect to a few risk factors associated with all accidents (speed, friction, lateral placement, following distance) these factors are ***not*** found to have less favourable values for the hazardous site than the matched comparison site	The hazardous road location is ***likely to be a false positive site***; that is the accidents are likely to be the result of random fluctuations mainly.

Kilde: TØI report 883/2007

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

2.8 Ranking of black spots and selection of treatments

Various methods have been proposed for ranking black spots for analysis and selection for treatment. The simplest is perhaps to rank sites identified as hazardous by the recorded number of accidents. However, what we are looking for are ways to reduce the accidents: it is not obvious that a more cost-effective treatment can be found at a site has recorded 10 accidents than at a site that has recorded 7 accidents.

Hauer et al (2004) discuss how best to rank hazardous road locations, which they refer to as sites with promise. They propose that the best ranking is one that assigns high rank to sites that can be treated cost-effectively, lower rank to sites that need more expensive treatments. Hence, ranking needs to incorporate information on the expected cost of treatments. Such information can usually only be obtained from a detailed engineering study, which is normally not performed until after sites have been ranked. Indeed, one of the purposes of ranking is to select those sites for which a more careful study is to be made in order to identify effective treatments. The task would therefore seem impossible: to rank sites according to the expected cost-effectiveness of treatment, one needs information that is normally not available until after a detailed engineering study has been performed, and no such study will be performed until sites have been ranked.

To get around this problem, Hauer et al (2004) first ranked sites by five different criteria. They then performed a detailed engineering study for 22 sites that were ranked high according to two of these five criteria: (1) The expected number of accidents (EB-estimate), (2) The expected number of severity-weighted accidents (EB-estimate). Once the detailed engineering study had been performed and estimates of the costs of treatment obtained for each sites, ranking by benefit-cost ratio was then compared to the original ranking. Sites were plotted in a diagram with cumulative costs of treatment on the abscissa and cumulative safety benefit (in monetary terms) on the ordinate. The various ranking criteria could then be compared in terms of how successfully they identified sites where the most cost-effective treatments could be introduced. It was concluded that the total expected number of accidents or the cost-weighted total expected number of accidents were best ranking criteria. In a sense, this confirms the finding reported above, that the EB-estimate of the number of accidents provides the best basis for identifying hazardous road locations.

Hauer et al (2004) rejected ranking hazardous road locations by the excess of their expected number of accidents above a reference level, a criterion analogous to the EB dispersion criterion discussed above. They argue that effective treatments are likely to reduce all accidents, not just those that represent an excessive number compared to what is normal for similar sites. This argument has clear merit as far as treatments affecting all accidents are concerned, like lowering speed limits. It is, however, less obvious that it is equally valid when treatments are directed at specific types of accidents. Thus, guard rails may prevent vehicles from going off

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

the road in a curve which is afflicted by this problem, but may not prevent rear-end collisions due to limited sight distance in the same curve.

An alternative ranking criterion is used in Norway (Statens vegvesen 2006). Any site that has recorded 4 or more accidents in 5 years is identified as a black spot. For each site, the annual cost of accidents is estimated, relying on the recorded number of accidents, but using mean cost rates for the various types of accident. Then, the normal annual cost of accidents for a similar site is estimated. A similar site is defined as a site of the same basic type (e.g. three leg junction) and having the same traffic volume as the black spot. The cost of accidents at the black spot are then compared to the normal cost of accidents at similar locations. The difference is taken and sites are ranked for detailed engineering analysis by the "excess" cost of accidents.

To account for the fact that treatments may not just influence the excessive number of accidents, but actually also reduce the normal number of accidents, the normal number of accidents and the costs associated with these accidents is multiplied by 0.8, implying that a 20% improvement of the normal level of safety is possible in addition to eliminating the excessive costs associated with an abnormally high number of accidents.

The attraction of this method for ranking sites for more detailed study is that it provides a benchmark for the maximum costs of treatment that can reasonably be incurred without making a project ineffective from an economic point of view (i.e. giving benefits smaller than costs). A possible weakness of the method is that it may not adjust sufficiently for regression-to-the-mean, by relying on the recorded number of accidents at black spots. On the other hand, it employs the average cost of accidents, thus avoiding the enormous inflation of cost estimates that would result if a fatal accident has been recorded at a black spot. It would, to a major extent, be the result of chance whether any of the accidents recorded at a black spot was a fatal accident.

2.9 Evaluation of the effectiveness of black spot treatment

Systematic evaluation of the effectiveness of black spot treatment is essential. For too long, the complexity of this task has been underestimated by researchers. As a result, a number of methodologically flawed evaluations have been made. A critical review of these evaluation studies is given by Elvik (1997). Some main points of his study will be presented.

Studies were classified according to whether or not they controlled for the following potentially confounding factors in before-and-after studies of black spot treatment:

1. Regression-to-the-mean
2. Changes in traffic volume
3. Long-term trends in the number of accidents

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

4. Accident migration, that is the tendency for accidents to "migrate" from treated black spots to other locations.

The classification was generous: studies that claimed to have controlled for any of the confounding factors were treated as having done so, although some studies did not explain in sufficient detail how they had controlled for the confounding factors.

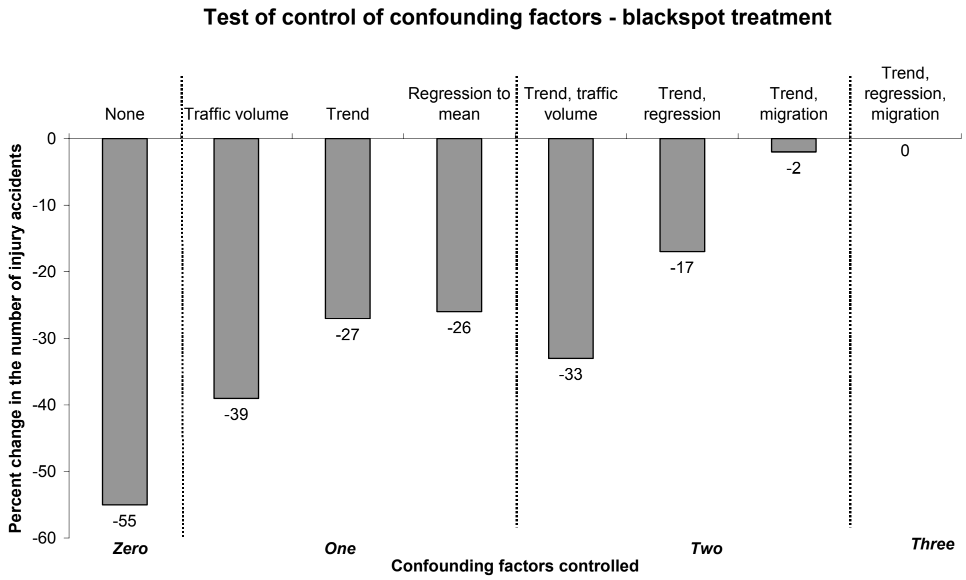


Figure 5: The importance of confounding factors in before-and-after studies of black spot treatment. Source: Elvik 1997

Figure 5 gives a sample of the results of the study. It shows the percentage change in the number of injury accidents attributed to black spot treatment, depending on which confounding factors studies controlled for.

In simple before-and-after studies that did not control for any of the four confounding factors, an impressive accident reduction of 55% was attributed to black spot treatment. In studies that controlled for regression-to-the-mean, long-term trends and accident migration, the effect attributed to black spot treatment was zero. There is a clear tendency in support of the Iron Law of Evaluation Studies: The more confounding factors a study controlled for, the smaller the effects attributed to black spot treatment.

Now, some people might wonder how we can know that a potentially confounding factor, say long-term trends, actually did confound a study. The answer is simple. If the effect attributed to the road safety measure differs depending on whether or not the potentially confounding factor is controlled for, then it does in fact confound study results. Potentially confounding factors do not, of course, always actually confound the results of a study. If there are no long-term trends in accidents, then this factor cannot confound. The point is that we cannot know whether or not a potentially confounding factor actually confounds a

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

study unless we control for it. The fact that a certain factor is potentially confounding is, in other words, a sufficient condition for trying to control for it.

Only an experimental study design in which units are assigned randomly to a treated and untreated group makes sure all potentially confounding factors are controlled for. In non-experimental studies, the best we can do is to control for the confounding factors that are known at any time, and for which relevant data can be obtained.

Let us return for a moment to Figure 5. It has been claimed that: "considerable safety benefits may accrue from application of appropriate road engineering or traffic management measures at hazardous road locations. Results from such applications at "black spots" demonstrating high returns from relatively low cost measures have been reported worldwide." (quoted from Elvik 1997). Is this claim justified? Take a careful look at Figure 5 and judge for yourself. The pattern shown in Figure 5 would seem to support a rather harsh verdict: The claim that black spot treatment is an effective way of preventing road accidents is totally unsubstantiated. It is based on an uncritical acceptance of studies that must be rejected because they did not control for important, and well known, confounding factors.

Today, the Empirical Bayes method is widely regarded as the "gold standard" for observational before-and-after studies of road safety measures. It should be recognised, however, that application of this method is not always straightforward and that if it is inappropriately applied, it can produce misleading results (Persaud and Lyon 2007). The fictitious data presented in Table 4 can be used to develop a simple example of the use of the Empirical Bayes method in a before-and-after study.

Suppose sites that recorded 4 or more accidents are regarded as black spots. There are 66 such locations. Further, suppose that 35 are selected for treatment, 31 are not. The 35 sites selected for treatment includes all 28 correct positives and 7 of the false positives. This assumption is reasonable, as even state-of-the-art techniques for accident analysis cannot guarantee that only correct positives are selected for treatment.

Sites selected for treatment had a total of 183 accidents before treatment, of which 148 at the correct positives and 35 at the false positives. The remaining false positives, not selected for treatment, had 151 accidents. In this data set, regression-to-the-mean is known (since accidents are assumed to be Poisson-distributed around the various mean values). The true long-term-expected number of accidents, after controlling for regression-to-the-mean, can be calculated to 112 for the correct positives (recorded 148), 21 for the false positives (recorded 35) and 90 for the false positives not selected for treatment (recorded 151).

It will be assumed that treatment is only effective for the correct positives, reducing their expected number of accidents by 25% (from 112 to 84). No effect is assumed for the false positives selected for treatment, nor for the other false positives.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 15 summarises the findings expected by various techniques for before-and-after studies.

Table 15: Results of hypothetical before-and-after study using various techniques

Technique	Recorded before	Expected after	Recorded after	Estimate of effect
Simple before-and-after	183	183	105	-43%
Before-and-after, non-treated as comparison	183	109	105	-4%
Empirical Bayes before-and-after	183	122	105	-14%
True situation	183	133	105	-21%

Kilde: TØI report 883/2007

The true situation, as determined from the hypothetical data, is a decline from a long-term expected number of accidents of 133 to 105, an accident reduction of 21%. If a simple before-and-after study is made, not controlling for regression-to-the-mean, the effect of the treatment will be considerably overstated (estimated to 43% versus the correct value of 21%). If the non-treated black spots are used as comparison group, the effect of the treatment will be underestimated. The ratio of after to before accidents in the comparison group (90/151) is used as a "control ratio" and multiplied by the recorded number of accidents before in the treated group (183), yielding an expected number of accidents of 109. This will over-adjust for regression-to-the-mean, as a stronger regression-to-the-mean effect is expected for the non-treated sites than for the treated sites.

The Empirical Bayes technique was implemented by predicting the expected number of accidents according to the following expression:

$$E(\lambda \mid x) = \left(\frac{\bar{X}}{\mathrm{Var}(X)}\right) \cdot \bar{X} + \left[1 - \left(\frac{\bar{X}}{\mathrm{Var}(X)}\right)\right] \cdot X \qquad (8)$$

This is the EB-estimate of the expected number of accidents applied to a population of sites for which no accident prediction model has been estimated (Hauer 1986). $\bar{X}$ is the mean number of accidents in a population of sites, Var(X) is the variance. Using the hypothetical data in Table 5, the mean is 0.779 and the variance is 2.003. The EB-estimate thus becomes.

[(0.779/2.003) · 0.779] + [(1 – (0.779/2.003)) · X]

In which X denotes the recorded number of accidents in the before-period. As can be seen from Table 15, the EB-method does slightly underestimate the true effect of the treatment, but it comes closer than any of the other techniques.

The source of the error in this case is that the EB-prediction is based on a considerably more heterogeneous population of sites than those that are considered for black spot treatment. This mean that the variance exceeds the mean and that the slope parameter of the EB-predictions (the expression within the first

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

brackets above) becomes too low. This, in turn, underlines the importance of basing EB-predictions on a reference group that is as similar to the treated group as possible (Persaud and Lyon 2007). The use of accident prediction models as part of the EB-method may help in this regard.

2.10 Key elements of state-of-the-art black spot management

The key elements of a state-of-the-art approach to black spot management can be summarised as follows:

1. Road accident black spots should be identified by reference to *a population of roadway elements for which the normal level of safety can be estimated*. Examples of populations of roadway elements include: sections of a specified length, curves with radius within a certain range, bridges, tunnels, three-leg junctions, four-leg junctions, etc. It should in principle be possible to enumerate all elements of each population.
2. Road accident black spots should be identified in terms of the expected number of accidents, i.e. *black spots are sites that have a higher expected number of accidents than the normal expected number for members of the population to which they belong*. Ideally speaking, a black spot should have a higher expected number of accidents than other similar locations due to specifically local risk factors. In practice, however, precise estimation of the contribution of local risk factors to accidents at black spots may not be possible.
3. The above points imply that *black spots cannot be reliably identified in terms of a critical count of accidents*. *For the purposes of accident analysis, it is nevertheless wise to identify only sites that have a certain minimum number of accidents as black spots*. The diagnostic performance of alternative critical values for the count of accidents should be tested and, if possible, the optimal value selected.
4. Inclusion of accident severity in the identification of black spots is best made by means of a preliminary analysis of accidents at black spots. For black spots with a given expected number of accidents, *the cost of the accidents should be estimated, preferably by relying on a model that allows estimation of the expected number of accidents at each level of severity*. In case such a model is not be applicable, the cost of accidents should be estimated on the basis of the recorded number of accidents. Sites that have a high mean cost per accident should be ranked high on a list for more detailed engineering analysis.
5. Detailed engineering analysis of black spots has two purposes: (1) To identify the true and false black spots and eliminate false black spots from further consideration. (2) To propose safety treatments for the true black

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

spots. *Accident analysis should be performed in two stages. The first stage is, by means of detailed examination of accidents, to suggest hypotheses regarding risk factors that may have contributed to the accidents. The second stage is to test the hypotheses developed in the first stage of analysis*. This can be done by means of a double blind comparison of the incidence of risk factors at each black spot and a comparison location with a good safety record.

6. *Treatments proposed for black spots should always be evaluated using state-of-the-art techniques. The Empirical Bayes method represents the current state of the art for before-and-after studies of road safety treatments*. It controls for (a) Local changes in traffic volume, (b) Long term trends in accidents, and (c) Regression-to-the-mean. If accident migration is an issue, an attempt to control for it should be made. Failure to control for all known confounding factors may result in grossly erroneous estimates of the effects of black spot treatment.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

3 Accident prediction models: a methodological review

Accident prediction models are an important element of state-of-the-art techniques both for black spot management and safety analysis of road networks. In view of this fact, it is important to develop good accident prediction models. This chapter discusses some difficulties that may arise in developing accident prediction models and provides some guidelines for evaluating accident prediction models from a methodological point of view. The following topics will be discussed:

1. Specification of a model
2. Choice of explanatory variables for inclusion in accident prediction models
3. Dual-state models
4. Multi-level models
5. Specification of functional relationships
6. Specification of residual terms
7. Evaluation of goodness of fit
8. The treatment of time
9. Controlling for endogeneity
10. Causal interpretation of relationships found
11. Assessing the predictive performance of a model
12. Application of model estimates in the empirical Bayes approach to road safety estimation
13. Assessing potential sources of error in predictive models

Each of the topics will be dealt with rather briefly. References are given to more extensive treatments.

A note on terminology is perhaps needed to clarify some key terms used in the discussion. The term accident prediction model usually denotes a multivariate model fitted to accident data in order to estimate the statistical relationship between the number of accidents and factors that are believed to be (causally) related to accident occurrence. The term "predictive" is somewhat misleading; "explanatory" would be a better term. Prediction refers to attempts to forecast

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

events that have not yet occurred, whereas accident prediction models are always fitted to historical data and can thus only describe, and perhaps explain, past events. The issue is not merely terminological; an example will be given of a model that almost perfectly reproduced the data it was fitted to, yet turned out to give badly wrong predictions for future years (Partyka 1991).

Terms that are used less frequently, but in the same sense as the term predictive model, include multivariate model, explanatory model or structural road accident model (Gaudry and Lassarre 2000).

3.1 Model specification

Logically speaking, the first decision made in developing an accident prediction model is the specification of the model. Model specification refers to the general formulation of a model in terms of the following characteristics:

1. Choice and definition of dependent variables.
2. Choice and definition of independent variables.
3. Specification of the functional form of the relationship between independent and dependent variables.
4. Specification of the residual terms of the model.
5. The treatment of time in a model.

For each of these items, a choice must be made between different alternatives. With respect to the dependent variable, the options include:

1. Using the number of accidents as the dependent variable.
2. Using the number of killed or injured road users as dependent variable(s).
3. Using an accident rate or injury rate as dependent variable.

The most commonly used dependent variable in accident prediction models is the number of accidents. If the number of injured road users is used as dependent variable, there may be problems of dependency between observations. Thus, if a road user was fatally injured in an accident, it is more likely that other road users involved in the same accident were also fatally or seriously injured. The use of multi-level models (to be discussed below) may be a way of dealing with this problem. Another option is to develop separate models for each level of injury severity.

Accident or injury rates (per million vehicle kilometres of travel) are rarely used as dependent variable in accident prediction models today. It has been recognised that the relationship between traffic volume and the number of accidents or injuries tends to be non-linear. Hence the use of a rate as a dependent variable is inappropriate, since it implies a linear relationship.

Points 2-5 on the list above are discussed in subsequent sections.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

3.2 Choice of explanatory variables

The expected number of accidents has traditionally been modelled as the product of exposure and accident rate:

$$\text{Expected number of accidents } (\lambda) = \text{Exposure (N)} \cdot \text{Accident rate (p)} \qquad (9)$$

Accident rate is traditionally defined as the (recorded) number of accidents per unit of exposure:

$$\text{Accident rate} = \frac{\text{Number of accidents}}{\text{Unit of exposure}} \qquad (10)$$

It is usually the recorded number of accidents that is used in the numerator when estimating the accident rate. This is equivalent to assuming that the recorded number of accidents is an unbiased estimate of the expected number of accidents, which is not always true. Exposure can be conceived of as the number of trials in a binomial experiment, and a commonly used unit of exposure is one kilometre of travel (Hakkert and Braimaister 2002). The idea, deeply rooted in probability theory, that the expected number of accidents depends on exposure and accident rate, is perhaps the source of the assumption traditionally made in road safety research that one can account for the effects of traffic volume on accidents by using accident rates. However, as will be discussed in the next section, this assumption is no longer tenable.

Based on this framework, the explanatory variables that are potentially relevant in accident prediction models can be placed in two main categories: (1) Variables describing exposure to accident risk, and (2) Risk factors that influence the number of accidents expected to occur per unit of exposure. Exposure has traditionally been treated as a unitary concept; in fact it is multidimensional. One may, for example, decompose exposure in terms of groups of road users or in terms of movements passing through an intersection. As will become clear, inadequate decomposition of exposure, or incomplete data on exposure, is a major weakness of current accident prediction models. Risk factors influencing accident occurrence, or the outcome of accidents, have long been recognised as very numerous and diverse.

Ideally speaking, the choice of explanatory variables to be included in an accident prediction model ought to be based on theory (Fridstrøm et al 1995). A theoretical basis for choosing explanatory variables might take the form of, for example, a causal model (Asher 1976), or path diagram, specifying the relevant variables and their paths of influence. In practice, a theoretical basis for identifying explanatory variables is rarely stated explicitly (Hauer 2004). The usual basis for choosing explanatory variables appears to be simply data availability. It is obvious that any analysis will be constrained by data availability. Nevertheless, the choice of explanatory variables should ideally speaking not be based on data availability exclusively. Explanatory variables should include variables that:

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

1. Have been found in previous studies to exert a major influence on the number of accidents,
2. Can be measured in a valid and reliable way,
3. Are not endogeneous, that is dependent on other explanatory variables included or on the dependent variable in the model.

In some cases, explanatory variables are entered stepwise into the model and included only if their relationship to the dependent variable is statistically significant. This procedure ensures that a parsimonious model is developed, i.e. a model that contains as few explanatory variables as possible. One should, however, not rely on statistical significance to decide whether or not a variable should be included. This may produce biased results.

Explanatory variables commonly included in accident prediction models fitted to data referring to roadway elements include:

- An indicator of exposure, often an estimate of vehicle kilometres of travel (usually including motor vehicles only)
- Variables describing the transport function of the road (motorway, main arterial, collector road, access road)
- Variables describing cross section (number of lanes, lane width, shoulder width, presence of a median, median width, etc)
- Variables describing traffic control (speed limit, type of traffic control at intersections)

Variables that are less often included in accident prediction models referring to roadway elements include:

- Variables describing alignment (horizontal and vertical curvature)
- Estimates of pedestrian and cyclist exposure
- Variables describing road user behaviour (speed, use of protective devices, etc)

Errors that may be caused by the omission of these variables, in particular incomplete exposure data are discussed in a later section.

3.3 Choice of model form

The basic form of nearly all modern accident prediction models is this (see e.g. Mountain et al 1996, Fridstrøm 1999, Gaudry and Lassarre 2000, Ragnøy, Christensen and Elvik 2002, Greibe 2003):

$$E(\lambda) = \alpha Q^{\beta} e^{\sum \gamma_i x_i} \qquad (11)$$

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

The estimated expected number of accidents, $E(\lambda)$, is a function of traffic volume, Q, and a set of risk factors, X_i (i = 1, 2, 3, … n). The effect of traffic volume on accidents is modelled in terms of an elasticity, that is a power, β, to which traffic volume is raised (Hauer 1995). This elasticity shows the percentage change of the expected number of accidents, which is associated with a 1 percent change in traffic volume. If the value of β is 1.0, the number of accidents is proportional to traffic volume, as traditionally assumed when using accident rates in road safety analysis. If the value of β is less than 1, the number of accidents increases by a smaller percentage than traffic volume. If the value of β is greater than 1, the number of accidents increases by a greater percentage than traffic volume. See e.g. Ivan (2004) for a discussion.

If road sections of varying lengths are used, an additional term may be added to the predictive equation to represent the effect of varying length of road segments.

The effects of various risk factors that influence the probability of accidents, given exposure, is generally modelled as an exponential function, that is as e (the base of natural logarithms) raised to a sum of the product of coefficients, γ_i, and values of the variables, x_i, denoting risk factors. Mathematically speaking, the following functions are identical:

$$\boldsymbol{Q}^{\boldsymbol{\beta}} = \boldsymbol{e}^{\ln(\boldsymbol{Q})\cdot\boldsymbol{\beta}} \tag{12}$$

This means that the common model form presented in equation 4 can be simplified to

$$E(\lambda) = e^{\sum \text{coefficients} \times \text{variables}} \tag{13}$$

This is in turn equivalent to a log-linear model:

$$\text{Ln}(E(\lambda)) = \alpha + \beta \cdot \ln(Q) + \gamma_i \cdot X_i \tag{14}$$

In models that refer to intersections, a fairly common model form is the following (see e.g. Brüde and Larsson 1993, Turner and Nicholson 1998, Miaou and Lord 2003, Persaud et al 2003):

$$E(\lambda) = \boldsymbol{\alpha}\, \boldsymbol{Q}_{MA}^{\boldsymbol{\beta}} \boldsymbol{Q}_{MI}^{\boldsymbol{\beta}} \boldsymbol{e}^{\sum \gamma_i x_i} \tag{15}$$

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

In this model, Q_{ma} refers to the number of vehicles entering an intersection from the major road, Q_{mi} refers to the number of vehicles entering an intersection from the minor road. Several versions of these basic model forms have been developed; see the papers of Turner and Nicholson (1998), Miaou and Lord (2003) and Oh, Washington and Choi (2004) for examples of alternative models.

There are few guidelines for the choice of model form. The choice of an exponential form is logical in view of the characteristics of the Poisson distribution. More generally, since $E(\lambda)$ cannot be zero or negative, multiplicative models are the only ones that are admissible. A multiplicative model does not need to be Poisson, however. Additive, linear models are rarely used today, as these models can give illogical results, like a negative predicted number of accidents. Interaction terms are not very common in accident prediction models, although one can easily imagine that, for example, a sudden change in both alignment and cross section may have a greater effect on accidents than the sum of the effects of changes in each of these design elements. Hauer (2004) offers some preliminary guidelines, arguing that accident prediction models should contain both a multiplicative and an additive portion:

Y = scale parameter × [(segment length for prediction) × (multiplicative portion) + (additive portion)]

The multiplicative part is intended to represent the effects of traffic volume and continuous hazards. The additive part is intended to represent the effects of point hazards, like driveways. Hauer, Council and Mohammedshah (2004) present an example of a model of this form. The choice of model form is closely related to the specification of functional relationships, discussed below.

3.4 Dual-state models

Accident prediction models differ in terms not just of the variables included, but also with respect to the assumptions made regarding the accident generating process. It is important to stress the fact that the accident generating process cannot be observed directly; only its outcome can be observed. The most common form of model is based on the assumption that accidents occur at a constant rate per unit of time in a given period. This rate will vary from place to place, and may vary from period to period. Models based on this assumption are referred to as single-state models.

Recently, an alternative family of models, usually referred to as zero-inflated Poisson or zero-inflated negative binomial models have been proposed (Shankar, Milton and Mannering 1997) and applied in some analyses (e.g. Lee and Mannering 2002). These models are based on the assumption that there are two modalities, or two states, for the accident generating process: a normal state, corresponding to the usual assumption of a constant expected number of accidents per unit of time, and a safe state, in which accidents will not occur. The resulting empirical probability distribution for the number of accidents will be a mixture of

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

a standard compound Poisson distribution (like the negative binomial distribution) and a distribution containing zero outcomes (i.e. no accidents recorded) only. The empirical distribution will then contain an excessive number of zeros compared to the standard negative binomial distribution.

The use of zero-inflated models has generated controversy. In a recent paper Lord, Washington and Ivan (2005) argue that the empirical basis of zero-inflated models is questionable. They show by means of simulation that an excess number of zeros can arise as a result of using small units in time and space for analysis (e.g. one year of accident data for 0.5 km sections, rather than, say, four years of accident data for 1 km sections) and when some of the units have very low exposure (an AADT of, say, less than 500) combined with a high accident rate. Neither of these sources of excess zeros can be attributed to a true dual-state process, in which one of the states is perfectly safe. Lord et al (2005) argue that the use of zero-inflated models represents a misplaced emphasis on finding models that fit the data perfectly, rather than models that make sense from a theoretical point of view.

A more plausible form of dual-state models, not represented in modern accident prediction models, was proposed more than forty years ago by Cresswell and Froggatt (1963). According to these models, accidents occur at a constant rate most of the time. There are, however, occasional "spells" in which the risk of accidents is temporarily increased. Cases of such spells, applied to road sections or intersections, include rainfall, unusually cold weather, or holiday seasons associated with increased travel. Other dual-state models have been described by, e.g. Lemaire (1995).

3.5 Multi-level models

Road accident data represent observations at three levels (Lenguerrand, Martin and Laumon 2006):

1. The accident level
2. The vehicle level
3. The road user level

In most accidents, more than one vehicle or road user is involved. This means that a hierarchical model should be applied to account for the fact that data at the accident level and vehicle level may vary less than data at the road user level. To see how this may occur, consider the hypothetical data given in Table 16.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 16: Hypothetical data to show the hierarchical nature of road accident data

Accident level	Vehicle level	Road user level
Head-on accident	Car 1	Male, 19 years, seat belt not worn, killed
Head-on accident	Car 1	Female, 18 years, seat belt worn, seriously injured
Head-on accident	Car 1	Female, 17 years, seat belt worn, slightly injured
Head-on accident	Car 2	Male, 38 years, seat belt worn, serious injured
Head-on accident	Car 2	Female, 34 years, seat belt worn, slightly injured
Head-on accident	Car 2	Child, 3 years, in child restraint, uninjured

Kilde: TØI report 883/2007

As can be seen, the data referring to road users vary from road user to road user with respect to at least three of the four variables included (age, gender, wearing of seat belts or child restraints and injury severity). However, data referring to vehicles involved vary only once, and accident data do not vary at all, since all six road users were involved in the same head-on accident. This means that the effects of the type of accident (head-on, rear-end, etc) on, for example, injury severity cannot be properly assessed unless the hierarchical nature of the data is appropriately modelled. The fact that all road users were involved in the same accident, and that some of the road users were occupants of the same vehicle means that the data referring to road users are correlated: the fact that one road user was seriously injured increases the probability that another road user was also seriously injured. Again, unless this is modelled correctly, the confidence intervals for the effects attributed to the higher level variables (i.e. the vehicle- and accident level variables) will become too small. Lenguerrand et al (2006) thus conclude: "Using the LM (logistic regression model assuming independent data) is theoretically false but in practice departures from the more appropriate and more complex models are minor."

Most accident prediction models use counts of accidents as the dependent variable. Except for data sets that include a significant number of secondary accidents (i.e. accidents that were caused primarily by another accident, for example, by way of road users becoming so distracted by an accident as to become involved in another accident), one may assume that accident data are statistically independent, in the sense that the occurrence of one accident does not influence the likelihood of another accident.

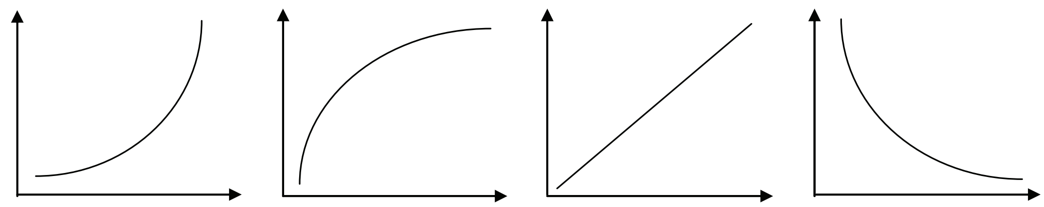
In a Norwegian model (Ragnøy, Christensen and Elvik 2002), the problem of correlations between levels of injury severity was circumvented by developing separate models for each level of injury severity.

3.6 Specification of functional relationships

Specification of functional relationships refers to the mathematical form of functions relating the number of accidents to one or more explanatory variables.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

In the standard formulation of accident prediction models, a power function is applied to describe the effects of exposure, and an exponential function applied to describe the effects of risk factors. These functional relationships can both take on many shapes, including all those shown in Figure 6.



Kilde: TØI report 883/2007

Figure 6: Functional forms consistent with a power model or an exponential model

These functional forms share one basic limitation. All the functions are monotonous throughout their range. This means that they do not permit "turning points", or local maxima or minima. More flexible functional forms can be fitted by applying Box-Cox transformations (Gaudry and Lassarre 2000) or by fitting polynomial functions by means of spline smoothing techniques (Miaou and Lord 2003). Hauer and Bamfo (1997) present an algorithm that can be used to explore the form of the function that best links an explanatory variable to a dependent variable. Thus, the development during the last ten years of powerful numerical algorithms that can be implemented on desktop computers means that virtually any imaginable functional form can be fitted to accident data. The challenge is therefore no longer to develop a model that may give an almost perfect fit to the data.

The requirement that a model be linear only means that it should be linear in parameters, not necessarily in functional form. There is nothing that prevents a linear model from providing non-monotonous effect curves. It is just a matter of including an appropriate set of variable transformations. In most cases, however, monotonicity is a sensible constraint to impose.

Examples of complex functional forms (Hauer et al 2004):

$$\frac{(\text{speed limit})^{3.991} e^{-3.991/30}}{\{50{,}000 \cup 150 \cup 2{,}000\}} \tag{16}$$

$$\left[e^{-0.171D} + 0.057D \right] \tag{17}$$

The first of these functions relates accidents to speed limit. The values listed in the denominator represent different types of road; in applying the function, one of these values (either 50,000, 150 or 2,000) is chosen. The second function relates

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

accidents to the degree of curve, a concept used in highway design in North America. A plot of the function shown in equation 12 is shown in Figure 7 (the function shown in equation 12 applies to PDO-accidents; similar functions were fitted for injury accidents and all accidents).

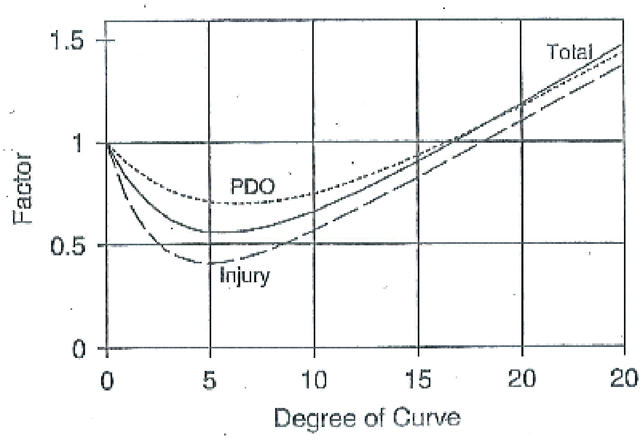


Figure 7: An example of a complex functional relationship. Source: Hauer, Council and Mohammedshah, 2004

3.7 Specification of residual terms

Observed variation in the number of accidents is nearly always a mixture of systematic and random variation. It is only the systematic part of the variation that can be explained by means of accident prediction models. A perfect model explains all systematic variation in accident counts. When estimating an accident prediction model, it is important to specify the distribution of the residual terms correctly. The residual term of a model is the part of systematic variation in accident counts, plus random variation, which is not explained by the model. If a model explains all the systematic variation in accident counts there is in a data set, the residuals will by definition contain random variation only and can be specified as Poisson distributed. Usually, however, a model will not be able to explain all systematic variation in accident counts. The residuals will then contain some over-dispersion, which can usually be adequately described by the negative binomial distribution.

As noted above, a number of probability distributions have been fitted to accident data. The Poisson and negative binomial are the most commonly used distributions; other distributions include zero-inflated Poisson or zero-inflated negative binomial (Shankar, Milton and Mannering 1997), dual-state Poisson (not zero-inflated), Poisson inverse Gaussian (Lemaire 1995), condensed negative binomial, (Morrison and Schmittlein 1981), the long distribution and the short distribution (Cresswell and Froggatt 1963). In principle, all these distributions may describe both the distribution of accidents in a population at risk and the distribution of the residual terms of an accident prediction model. For further

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

discussion, see Fridstrøm et al (1995), Turner and Nicholson (1998), Hauer (2001), Miaou and Lord (2003) and Elvik (2004).

3.8 Evaluation of goodness of fit

Several measures have been proposed to evaluate the goodness of fit of accident prediction models. Miaou (1996) shows that the squared multiple correlation coefficient is not suited as a measure of the goodness of fit of an accident prediction model. Maher and Summersgill (1996) discuss the use of the scaled deviance and the log-likelihood ratio as measures of goodness-of-fit, concluding that the scaled deviance is not suited for data sets that have a low mean number of accidents. Fridstrøm et al (1995) discuss and compare five measures of goodness-of-fit. One of these, termed the Elvik index, is derived from the over-dispersion parameter of negative binomial accident prediction models and will be discussed in more detail, because it is computationally simple and has a number of the desirable characteristics for a measure of goodness-of-fit listed by Miaou (1996).

Recall that the total variation in the count of accidents found in a sample of study units can be decomposed into random variation and systematic variation (Hauer 1997):

$$\text{Total variation} = \text{Random variation} + \text{Systematic variation} \tag{18}$$

There is systematic variation in number of accidents whenever the variance exceeds the mean. This is referred to as over-dispersion. The amount of over-dispersion found in a data set, can be described in terms of the over-dispersion parameter, μ, which is defined as follows:

$$Var(x) = \lambda \cdot (1 + \mu\lambda) \tag{19}$$

Solving this with respect to the over-dispersion parameter (μ) gives:

$$\mu = \frac{\frac{Var(x)}{\lambda} - 1}{\lambda} \tag{20}$$

The success of a model in explaining accidents can be evaluated by comparing the over-dispersion parameter of a fitted model to the over-dispersion parameter in the original data set. Table 17 presents a data set for national highways in Norway, showing the number of fatalities per kilometre of road during 1993-2000 (Ragnøy, Christensen and Elvik 2002). The mean number of road accident fatalities per kilometre of road was 0.0646; variance was 0.0976. The over-dispersion parameter can be estimated to 7.91. A multivariate accident model was fitted to the data, assuming a negative binomial distribution for the residuals. The coefficients estimated for this model are listed in table 17. The over-dispersion parameter for the model was 2.39. Inserting this into equation 9 gives an

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

estimated variance of 0.0745. The contributions of various factors to the observed variance in accident counts can be determined as follows:

Random variation = 0.0646/0.0976 = 0.662 = 66.2% of all variance

Systematic variation = (0.0976 – 0.0646)/0.0976 = 0.338 = 33.8%

Systematic variation explained by model = (0.0976 – 0.0745)/0.0976 = 0.237 = 23.7%

Systematic variation not explained by model = (0.0745 – 0.0646)/0.0976 = 0.101 = 10.1%

The model explains 0.237/0.338 = 0.701 = 70.1% of all systematic variation found in this data set.

Most models assume a constant over-dispersion parameter. If, however, the “size” of the units of analysis differs (road sections with different lengths, for example), it is more correct to treat the over-dispersion parameter as a variable (Hauer 2001) and model it as a function of, for example, section length.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 17: Number of road accident fatalities per kilometre of road, national highways, Norway 1993-2000. Source: Ragnøy, Christensen and Elvik 2002

	Distribution of road sections by number of fatalities			Accident prediction model	
Number of fatalities	**Actual**	**Poisson**	**Negative binomial**	**Explanatory variables**	**Coefficient**
0	19957	19728	19974	Constant	-7.154
1	895	1274	854	Ln(AADT)	0.842
2	135	41	163	Speed limit 50 km/h	Reference category
3	43	1	39	Speed limit 60 km/h	-0.020
4	9	0	10	Speed limit 70 km/h	0.385
5	3	0	3	Speed limit 80 km/h	0.172
6	1	0	1	Speed limit 90 km/h, rural road	0.090
7	0	0	0	Speed limit 90 km/h, class B road	0.610
8	1	0	0	Speed limit 90 km/h, class A road	0.879
N	21044	21044	21044	Number of lanes	-1.967
				Number of intersections/km	0.082
				Dummy for trunk road	0.255
Mean	0.0646				
Variance	0.0976			Estimated variance	0.0745
Over-dispersion parameter	7.91			Over-dispersion parameter	2.39

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

3.9 The treatment of time

Time is usually not entered explicitly as a variable in accident prediction models. Some accident prediction models are based on data that refer to several years and may want to describe changes over time explicitly. Different approaches have been taken to this. One option, exemplified by Hauer et al (2002B), is to develop a separate accident model for each period concerned. The drawback of this option is that the number of accidents for a single period will be smaller than for multiple periods, making a reliable estimation of model parameters more difficult. Sometimes, inexplicable fluctuations in model parameters will be found, which raises the issue of whether smoothed estimates should be applied.

A second option, exemplified in the models developed by Mountain et al (1998) is to account for changes over time in terms of a trend parameter. The coefficients for the explanatory variables, except for time, then apply to the entire period covered. The trend term is used to adjust the predicted numbers for year-by-year changes. There are two limitations to this approach: The first one is that, usually, a single term will be estimated to represent changes over time. This means that these changes are assumed to be constant, like a constant percentage decline or increase in the number of accidents from one year to the next. Long-term trends may, however, not always be constant. Another limitation is that in estimating a trend term, the assumption is normally made that accident data for successive years are independent, which is not the case. Time series of accident data tend to display autocorrelation, that is successive observations are correlated. Unless this is accounted for, the standard error of the trend term will be underestimated.

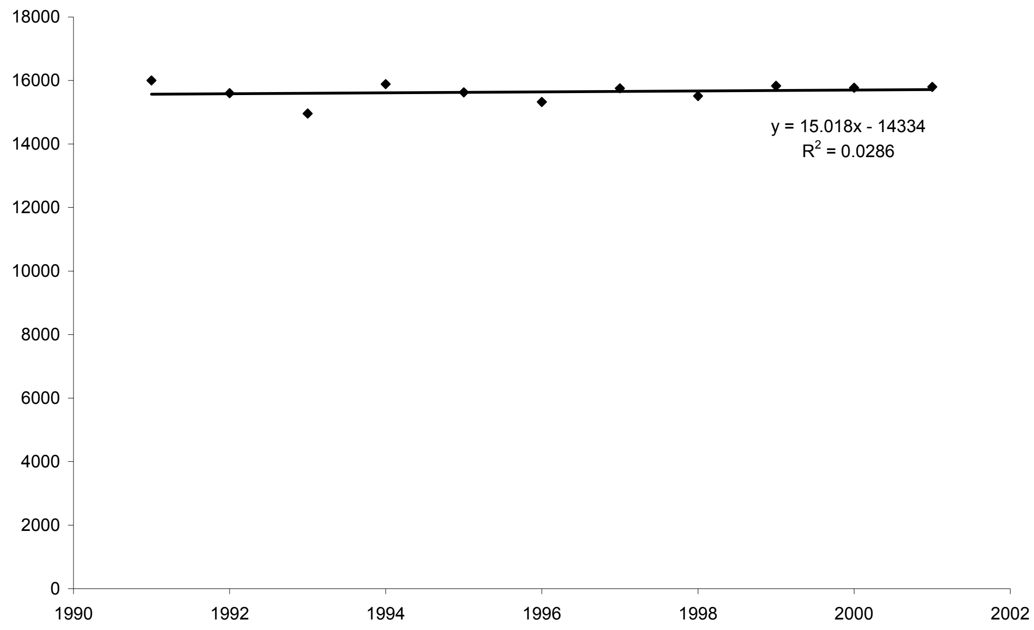
Successive observations of a Poisson variable are, however, uncorrelated. If an accident prediction model has been fitted that accounts for all systematic variation in accident counts, the residuals will not display autocorrelation and not cause erroneous estimates of standard errors.

A third approach, advocated by Lord and Persaud (2000), is to apply a generalised estimating equations procedure in order to explicitly account for the autocorrelation present in accident data and estimate separate adjusting factors for each year covered by the model. This is statistically more complex than the other two approaches, but is theoretically more correct.

An issue related to the treatment of time in accident prediction models is the question of how long the period covered by the data can be without having to enter time as an explanatory variable in the model. Consider the data shown in Figure 8 referring to injury accidents in Sweden from 1991 to 2001. The stability of the figures is striking. No clear trend can be discerned.

Nevertheless, factors influencing the number of accidents may have changed during this period. It is likely that traffic volume has increased, while some of the factors affecting accident rate have influenced it favourably, leading to a very stable total number of accidents. If mean values applying to the entire period are used for these explanatory variables, important changes in their effects on accidents may go undetected. For a further discussion of this point, see the discussion on the use of averages for traffic volume in section 3.14.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961



Kilde: TØI report 883/2007

Figure 8: Count of police-reported injury accidents in Sweden 1991-2001

3.10 Controlling for endogeneity

Endogeneity can be a problem in accident prediction models, in particular models that include road safety treatments among the explanatory variables. Endogeneity refers to the tendency for the dependent variable in a model – typically the count of accidents – to influence one or more of the independent variables used to explain the dependent variable. The presence of endogeneity in an accident prediction model can lead to grossly erroneous estimates of the effects of some variables.

In an instructive paper, Kim and Washington (2006) explain what endogeneity is, how it may affect model estimates, and how to control for it. Consider, as an example, the use of left-turn lanes in junctions. Left-turn lanes are more likely to be installed in junctions that have many vehicles turning left, and many accidents associated with left turns, than in otherwise similar junctions that do not experience these problems. In short, the selection for treatment is strongly influenced by the dependent variable of the model, i.e. the count of accidents. If this is not recognised, one may erroneously find that the presence of left-turn lanes is associated with a higher expected number of accidents than if there are no left turn lanes.

The solution to the problem is to employ two-stage modelling. A model is first developed to describe the selection for treatment by means of left turn lanes, i.e a model that explains the probability that a junction will have left turn lanes. The results of that model are then used in the main model, which will thus control for endogeneity. This can make a great difference to the results, as shown in Table 18. The table is based on Tables 2 and 4 in the paper of Kim and Washington.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 18: Estimated coefficients for selected variables without and with control for endogeneity. Source: Kim and Washington 2006

Coefficients for selected variables	Without control for endogeneity	With control for endogeneity
Log of AADT major road	0.3227	0.1289
Log of AADT minor road	0.2868	0.2756
Left-turn lane indicator	0.6897	-0.1653
Number of driveways	0.1142	0.1583
Lighting indicator	-0.5925	-0.6520

It is seen that without control for endogeneity, left turn lanes are found to be associated with an increased number of accidents. With control for endogeneity, left turn lanes are found to improve safety. The coefficients for the other variables did not change materially as endogeneity was controlled for.

3.11 Causal interpretation of relationships found

Discussions about the causality of statistical relationships usually start with the statement: Correlation does not equal causation. Correlation is necessary for causation, but not sufficient. What more should there be to a statistical relationship in order to interpret it as evidence of a causal relationship? This question has been discussed at some length by Elvik (2001, 2007) and Hauer (2005A), among many others. It is a complex question; space does not permit an extensive discussion of it in this paper. Suffice it to note that the following criteria (the list is not necessarily exhaustive) have been proposed to help assess if a statistical relationship is causal:

1. Internal consistency of the relationship, with respect to, for example, subsets of data in a study or different specifications of multivariate models.
2. Invariance with respect to potentially confounding factors, meaning that a relationship does not vanish when potentially confounding factors are controlled for. Application of this criterion requires a clear definition of what a potentially confounding variable is (see below).
3. Plausibility in terms of a known mechanism or well-established scientific law that accounts for the statistical relationship between cause and effect.
4. Support for counterfactual statements, meaning that the relationship has a genuine predictive capacity (see example below).

The first criterion, internal consistency, means that the coefficient estimated for a variable should remain substantively unchanged (i.e. identical within the bounds of statistical uncertainty) across different model specifications or subsets of the data.

This criterion is closely related to the second criterion, which is that to defend a causal interpretation of a coefficient showing a statistical relationship, this relationship should not vanish when potentially confounding variables are controlled for. One should, however, take great care to specify a model of the

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

relationships between variables that clearly identifies those variables that are potentially confounding and those that are not. Unless statistical estimation is guided by such a model, there is a risk of controlling for variables that should not be controlled for, and neglect to control for variables that should be controlled for.

Findings emerging from accident prediction models can be more or less plausible. It is, for example, not very plausible that raising the speed limit will improve the safety of a road. Yet, a statistical association suggesting that this is the case has been found in many models (see e.g. Ragnøy et al 2002, Hauer et al 2004).

Ideally speaking, model building seeks to reproduce law-like relationships. A scientific law supports counterfactual statements (see next section). This criterion will be satisfied, or at least to some extent supported, if an accident prediction model produces broadly correct predictions for a data set that was not used in fitting the model. Such predictive success indicates that the relationships included in a model are valid in general, and not just idiosyncratic or local.

3.12 Assessing the predictive performance of a model

As noted before, to predict is not the same as to explain. Most accident prediction models are not in fact predictive models, but explanatory models. Their predictive performance is mostly unknown. One of the very few models whose predictive performance has actually been tested, is a model developed by Partyka. The predictive performance of this model is shown in Figure 9 (Partyka 1991). The original model was a very simple model fitted to fatality count data for the United States from 1960 through 1982. The model included just three numerical variables and a dummy variable for the year 1974. As shown by Figure 9, the fitted values trace the actual values as a shadow; in fact the two series (actual and predicted) are almost indistinguishable. The squared multiple correlation coefficient for the model fitted to 1960-1982 data was 0.98 (leaving aside for the moment the question of whether this is a good measure of goodness-of-fit).

But look what happened when this model was used to predict for the years 1983-1989. By 1989, it over predicted the number of fatalities by nearly 20,000. The fact that this model fitted past data as perfectly as any model possibly could, did not ensure that it predicted correctly. Explaining past trends does not ensure that future trends can be reliably predicted.

There are at least two ways of evaluating the predictive performance of an accident prediction model. The first one is to use the model to predict accident counts in future years, as shown above. The second one is to use only half the data set to fit a model and use the other half of the data set to test its predictive performance. One rarely sees any of these tests of predictive performance applied.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

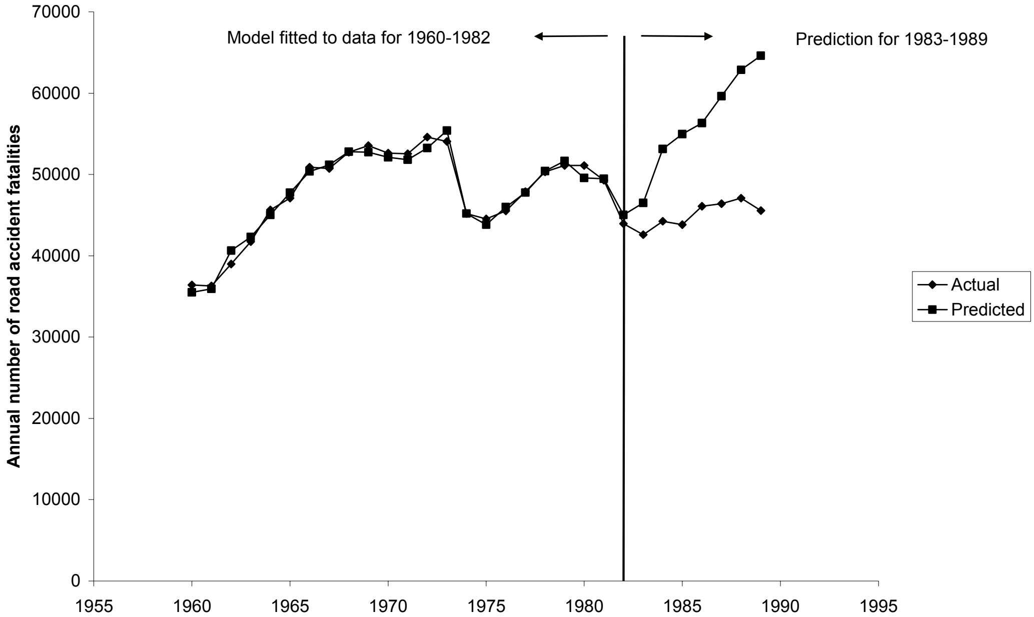


Figure 9: Predictive performance of a model fitted to fatality data for the United States. Based on Partyka 1991.

Testing predictive performance is essential if one wants to support a causal interpretation of model estimates. One of the criteria of causality is that statistical relationships are reproduced over time in different contexts. In fact, to interpret a statistical relationship as a statement of a scientific law, the relationship must support counterfactual statements (Hempel 1965). A counterfactual statement is an empirically testable statement about what would have happened if a certain causal factor had not been present. A trivial example will clarify this idea.

We observe that the water in a small pond has frozen during the night. From the meteorological service, we are able to verify that the temperature dropped below zero degrees Celsius for sufficiently long last night for the water to freeze. We are then able to explain why the water was frozen in the morning. A counterfactual statement would then be: If the temperature had not dropped below zero degrees Celsius, the water in the pond would not have frozen. This is contrary to what actually happened; yet it is a testable statement. The statement: Water freezes only when the temperature drops below zero, is a scientific law, since it can be verified that it supports the above counterfactual statement.

By the same token, if the number of accidents increases when traffic volume increases, we would expect the number of accidents *not* to increase when traffic volume does *not* increase. This, like any other statement about the association between accidents and explanatory factors, is of course only valid when everything else remains constant. The assumption that everything else remains constant is never correct in observational road safety studies. Partyka (1991), for example, speculates that the poor predictive performance of her model may be attributed to an unforeseen decline in drinking and driving in the United States in the nineteen eighties. Does this mean that the statistical relationships fitted in the original model broke down after 1982, and did not have the capacity to support counterfactual statements? There is no way of observing the counterfactual

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

condition directly; we only know what actually happened, not what might have happened. Yet, it is nevertheless possible to assess what would have happened if the decline in drinking and driving had not occurred. This can be done by refitting the model to data for the period 1960-1989, adding the years 1983-1989 to the data set. If the coefficients estimated in the original model remain substantively unchanged in the refitted model, this indicates that the effects captured by these coefficients are stable over time and invariant with respect to the presence or absence of other causal factors influencing fatalities. Such stability and invariance indicates a support for counterfactuals, albeit much weaker than the support usually obtained in the natural sciences.

In Partyka's original model, the coefficient for the number of unemployed workers was –1.8569. In the revised model, the coefficient was –1.3686. In the original model, the coefficient for the non-labour force was +0.9616. In the revised model, it was +1.2934. Standard errors are not given, but the sign and magnitude of the coefficients are similar, suggesting that effects are broadly speaking invariant and unaffected by the decline in drinking and driving. This suggests that the effects are causal.

3.13 Application of model estimates in the empirical Bayes approach to road safety estimation

An accident prediction model gives an estimate of the expected number of accidents for a roadway element that has a certain combination of traits. In most models, these include traffic volume, characteristics of highway geometry and type of traffic control. Most accident prediction models will not include all factors that produce systematic variation in accident counts. Hence, estimates of the expected number of accidents derived from accident prediction models are mean values for units that have a given combination of traits. The expected number of accidents for a specific unit will normally differ from the mean value for units that have similar general traits.

What is the best estimate of the long term expected number of accidents or accident victims for a given roadway element, given the fact that we know some, but not all of the factors affecting accident occurrence? According to the empirical Bayes method (Hauer 1997), the best estimate of safety is obtained by combining two sources of information: (1) The accident record for a given site, and (2) An accident prediction model, showing how various factors affect accident occurrence. Denote by R the recorded number of accidents, and by λ the normal, expected number of accidents as estimated by an accident prediction model. The best estimate of the expected number of accidents for a given site is then:

$$E(\lambda|r) = \alpha \cdot \lambda + (1 - \alpha) \cdot r \qquad (21)$$

The parameter α determines the weight given to the estimated normal number of accidents for similar sites when combining it with the recorded number of

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

accidents in order to estimate the expected number of accidents for a particular site. The best estimate of α is:

$$\alpha = \frac{1}{1+\frac{\lambda}{k}} \qquad (22)$$

λ is the normal expected number of accidents for this site, estimated by means of an accident prediction model and k is the inverse value of the over-dispersion parameter of this function, that is $1/\mu$. To illustrate the use of the empirical Bayes method, suppose that the normal expected number of accidents for a 1-kilometre road section during a period of eight years has been estimated by means of an accident prediction model to be 3.73. The over-dispersion parameter for this model is 0.3345; hence k is 2.99. The weight to be given to the estimate based on the prediction model thus becomes 1/[1 + (3.73/2.99)] = 0.445. Seven accidents were recorded. The long term expected number of accidents is estimated as:

$E(\lambda|r) = 0.445 \cdot 3.73 + (1 - 0.445) \cdot 7 = 5.54.$

The interpretation of the three different estimates of safety can be explained as follows. 3.73 is the number of accidents one would normally expect to occur at a similar site, that is one which has the same traffic volume, the same speed limit, the same number of lanes, etc, as the site we are considering. 7 accidents were recorded. Part of the difference between the recorded and normal number of accidents for this type of site is due to random variation. An abnormally high number of accidents due to chance cannot be expected to continue; a certain regression-to-the-mean must be expected. In the example given above, the regression-to-the-mean expected to occur in a subsequent eight year period is (7 – 5.54)/7 = 0.209 = 20.9%. The difference between the site-specific expected number of accidents (5.54) and the normal, expected number of accidents for similar sites (3.73) can be interpreted as an effect of local risk factors for the site, causing it to have a higher expected number of accidents than similar sites.

Applying the empirical Bayes method to a site that had 0 recorded accidents, but was otherwise identical to the site used as an example above, gives a site-specific expected number of accidents during eight years of 1.66 (the normal, expected number of accidents for similar sites was 3.73). For this site, the difference between the site-specific expected number of accidents and the normal, expected number of accidents for similar sites can be interpreted as the effect of local safety factors, which are factors causing the site to be safer than otherwise similar sites.

3.14 Assessing potential sources of error in predictive models

There are many sources of error in accident prediction models. The most frequently discussed sources of error include:

- Omitted variable bias
- Errors due to co-linearity among explanatory variables

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

- Wrong functional form for relationships between variables

To illustrate these potential problems, examples will be given. Brüde and Larsson (1993) fitted the following rather simple accident prediction models to data for intersections in Sweden:

Number of pedestrian accidents = 0.0000734 x $MV^{0.50}$ x $PED^{0.72}$

Number of bicycle accidents = 0.0000180 x $MV^{0.52}$ x $CYC^{0.65}$

MV is the number of motor vehicles (AADT = annual average daily traffic), PED is pedestrian volume, and CYC is cyclist volume.

Based on these functions, the number of accidents can be estimated for any combination of values for the number of motor vehicles and the numbers of pedestrians or cyclists.

These functions suggest a log-linear relationship between exposure and the number of accidents. If, as an example, the number of pedestrians increases from 500 to 1,000, and the number of motor vehicles increases from 5,000 to 10,000, the number of pedestrian accidents (that is accidents in which pedestrians are struck by cars) increases by a factor of nearly 2.33. In other words, the number of accidents is more than doubled when total traffic volume is doubled (from 5,500 to 11,000). Despite this, the risk run by each pedestrian, at a given amount of motor traffic, declines strongly as the number of pedestrians increases. If the number of pedestrians increases from 100 to 1,000, the risk of getting injured, stated as the number of pedestrian accidents per pedestrian exposed, drops by about 50%. A further increase in the number of pedestrians from 1,000 to 2,000 is associated with a further reduction in the injury rate per pedestrian of some 17%.

Now suppose pedestrian volume had not been known and a model predicting pedestrian accidents had been fitted to data containing motor vehicle volume as the only explanatory variable. The exponent for motor vehicle volume then becomes 0.90, as opposed to 0.50 in the model that included data on pedestrian volume in addition to motor vehicle volume. Because motor vehicle volume and pedestrian volume are correlated, the coefficient for motor vehicle volume will contain part of the effect of pedestrian volume when that is not included in the model. This is an example of omitted variable bias. The coefficient for motor vehicle volume is biased, because it includes part of the effect of pedestrian volume as well as the effect of motor vehicle volume.

Jonsson (2005) compared model coefficients for models that included bicyclist or pedestrian volume and models that did not include these variables. In the model that included both motor vehicle and bicyclist volume, the coefficient for motor vehicles was 0.76 and the coefficient for bicyclist volume was 0.35. When bicyclist volume was omitted, the coefficient for motor vehicle volume changed to 0.93. Similarly, for pedestrian volume, the coefficients when it was included were 0.83 for motor vehicle volume and 0.38 for pedestrian volume. When pedestrian volume was omitted, the coefficient for motor vehicle volume changed to 0.92.

How can we know if a model is afflicted by omitted variable bias? The answer is that we can never know this for certain. Even a model that has a very high explanatory power may be biased due to omitted variables, since any omitted

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

variables could be correlated both with the variables included in the model and the residual term of the model.

Possibly the most common form of omitted variable bias in current accident prediction models is the incompleteness of exposure data. These data rarely include pedestrian or cyclist exposure.

Explanatory variables in accident prediction models tend to be correlated, sometimes to such a high degree that inclusion of both or all the correlated variables may lead to imprecise estimates of the coefficients. Estimates are not biased, but will be associated with large standard errors. A case in point is the study of quantified road safety targets by Elvik (2001). In that study, each country was identified by a dummy variable. One of the countries included was the United States. Inclusion of the United States caused problems in the multivariate analysis, because the dummy variable identifying the United States was almost perfectly correlated with the fatality count, see figure 10.

The correlation between these two variables is 0.989, that is virtually 1. It is therefore nearly impossible to estimate the effects on fatality counts of any other variable very precisely in analyses that include the United States, but exclude measures of travel exposure (the inclusion of which might alleviate the problem, as travel exposure is very much greater in the United States than in the other countries, thus removing the effect of "size" of the country).

Kilde: TØI report 883/2007

Figure 10: Correlation between country dummy for United States and fatality count

There is little guidance concerning the choice of functional form in accident prediction models, although Hauer and Bamfo (1997) and Hauer (2004) give some useful hints. An extensive discussion is also given by Fridstrøm (1999), who relies on Box-Cox transformations of variables in order to identify the most

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

correct functional form. The use of such transformations allows for great flexibility and does not restrict the choice of functional form to any specific form, such as logarithmic, exponential, etc. Mensah and Hauer (1998) discuss two problems related to the use of average values when estimating the relationship between traffic volume and accidents. Both these averaging problems can lead to the fitting of wrong functional forms for the relationship between traffic volume and accidents.

The first problem is called "argument averaging". It occurs when traffic volume is represented by an average value, like AADT, rather than the actual traffic volume at the time of each accident. Traffic volume is known to vary during the day, the week and from month to month. To correctly describe the relationship between traffic volume and accidents, data reproducing these variations, rather than averaging them out, should be used in fitting models. Mensah and Hauer show that using an average value can lead to a biased functional form.

The second problem is called "function averaging". It occurs when a single function is used to model the relationship between traffic volume and accidents, when there is reason to believe that this relationship varies, depending on circumstances. As an example, traffic volume in darkness may have a different relationship to accidents than traffic volume during daytime. Separate functions should be fitted for daytime and darkness to reflect this fact. When a single function is fitted, representing all 24 hours of a day, it may misrepresent the relationship between traffic volume and accidents. Predictions based on a single function will not necessarily be identical to the mean of predictions based on separate functions, in particular not if these functions are all non-linear. The bias could be substantial.

In principle, the problem could be solved, not by developing separate functions for daylight and darkness, but by including light conditions as an explanatory variable in a model.

3.15 Concluding remarks: criteria for assessing the quality of accident prediction models

Despite the many problems identified in this section, there is no doubt that the development of what we may term "modern" accident prediction models during the past 15 years represents a major step forward in road safety research. Road safety research is now rapidly becoming a mature scientific discipline (Hauer 2005B), a discipline that can be taught in universities and that provides a basis for a rational approach to road safety management.

Development in the field of accident modelling has been so rapid, that some models that were considered as state-of-the-art only ten years ago look somewhat primitive today. There is today a danger – clearly pointed out by Lord, Washington and Ivan (2005) – of moving too far in the direction of mathematical sophistication and perfect fitting of models. Accidents are a very complex phenomenon; hence models also need to be complex in order to faithfully reproduce the main features of reality. Yet, the art of model building is, and will always be, the art of making the right simplifications. A good model is not necessarily an immensely complex model that perfectly fits the data in every

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

detail. A good model is rather the simplest possible model that adequately fits the data, and that contains relationships that may be presumed to hold in general, and not be an idiosyncratic feature of a particular data set or an esoteric model formulation.

Based on the discussion in this section, the following criteria are proposed for assessing the quality of accident prediction models:

1. As a basis for developing a model, *the probability distribution of accidents in the original data set should be investigated*. This investigation should include several of the most commonly used probability distributions for accidents. It is important to note, however, (see example in next chapter) that the initial distribution will not necessarily conform to any known probability distribution for accidents.
2. The residual terms of the model should be specified. A negative binomial distribution of residuals is often a reasonable hypothesis. *The structure of residuals should always be tested*.
3. *Separate models should be developed for accidents at different levels of severity*. As a minimum, separate models are required for fatal accidents, injury accidents (sometimes including fatal accidents) and property-damage-only accidents. Alternatively, multi levels models can be used.
4. *Separate models should be developed for different types of roadway elements*. Roadway elements include: road sections, intersections, bridges, tunnels, curves, railroad-highway grade crossings.
5. *Data on exposure should be decomposed to the maximum extent possible*. For road sections, these data should preferably indicate the proportions of all traffic made up by heavy vehicles, cars, motorised two-wheelers, pedestrians and cyclists. For intersections, exposure should be specified according to traffic movements passing the intersection.
6. *The functional form used to describe the relationship between each independent variable and accidents should be explicitly chosen* and reasons given for the choice. Alternative functional forms should be tested as a basis for the choice made.
7. *Explanatory variables should be entered stepwise into the model*. Variables describing exposure should always be entered first. When presenting the model, the full array of coefficients estimated at each stage should be presented, to allow an examination of the stability of the coefficients with respect to which variables were included in the model.
8. *The correlations between explanatory variables should be examined* to detect the possible presence of co-linearity. There is, however, no good solution to the problem of co-linearity. It does not bias coefficient estimate, but makes them highly uncertain.
9. *The overall goodness-of-fit of the final model should be reported* in a way that permits variation in accident counts to be decomposed into: (a) Systematic variation explained by the model, (b) Systematic variation not explained by the model, (c) Random variation.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

10. *The structure of any systematic variation not explained by a model should be examined* and a choice made as to whether over-dispersion is adequately described by a single parameter or should be modelled by a variable parameter.
11. *Any model should explicitly identify those variables for which a causal interpretation is sought* and those variables that are to be considered as confounding with respect to the causal relationships evaluated.
12. *Explicit operational criteria for causality should be stated in models seeking causal interpretation of their findings*. By operational criteria are meant criteria that can be evaluated empirically. Causal interpretations should only be proposed if all important operational criteria are met. Possibly the best way to test for causality is out-of-sample predictions.
13. *The possible presence of omitted variable bias should always be discussed*. It is understood that no accident prediction model can be "complete" by including absolutely every conceivable variable that may influence accident occurrence.
14. *The predictive performance of an accident prediction model should be tested*. This is done by applying the model to a data set that was not used in developing the model.
15. *Accident prediction model should permit results to be synthesised*. This means that any accident prediction model should report the standard errors of all coefficients in such a way as to permit a formal synthesis of the findings of multiple accident prediction models (meta-analysis).

These criteria can be further developed into a quality scoring system for accident prediction models, designed to assign a numerical quality score to each model. This quality score will be an important piece of information when synthesising the findings of several accident prediction models.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

4 Safety analysis of road networks

This chapter will discuss network safety management, or safety analysis of road networks. As many of the conceptual issues are analogous to those that arise in black spot management – such as trying to account for sources of systematic variation in accident counts and controlling for effects of random variation (regression-to-the-mean), the conceptual discussion will not be repeated here. Three currently used systems for network safety management will be presented. Before presenting these systems, a short overview of some choices that must be made when developing safety analysis of road networks as a tool for network safety management will be discussed.

4.1 Stages of safety analysis of road networks

4.1.1 Determining the scope and level of analysis

The scope of a road network safety analysis is usually the entire road system of a jurisdiction. A jurisdiction may, for smaller states, include the whole country (Norway), or it may include a federal state (Germany, United States). Sometimes analysis is done by route, usually applying an official route numbering system. Whichever approach is taken, a network safety analysis will usually comprise at least several hundred, more commonly perhaps several thousands, of kilometres of road. Traffic volume and highway design parameters will normally display great variation among the roads selected for analysis.

In most countries, roads that are classified and numbered will have numbered sections. Each of these sections is, ideally speaking, homogeneous with respect to traffic volume and other factors influencing the number of accidents, but sections may differ greatly among themselves with respect to length, traffic volume and other factors influencing the number of accidents. A choice must be made on how to define the "elementary unit" for analysis. Each such elementary unit ought, ideally speaking, to be homogeneous (i.e. each independent variable should take on the same value throughout the unit, e.g. a section should not have two different speed limits, one for half the length, another for the other half) with respect to factors influencing the number of accidents. This will sometimes require the use of rather short road sections as elementary units. A disadvantage of using short road sections is that it greatly reduces variation in the number of accidents. If, as an example, each section is just 10 metres long, almost all sections will have 0 accidents and some very few will have 1 accident. The prospect of reliably identifying factors explaining variation in accident counts is then greatly diminished, as nearly all observed variation will be random. Furthermore, some of the causal factors for accidents exert their influence over a stretch of road at least as long as the braking distance.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

On the other hand, by using very long road sections one ironically runs into the same problem, but from the opposite end of the spectrum. The use of very long road sections necessitates the smoothing of data pertaining to shorter subsections, for example by using mean values for traffic volume for a long road section along which traffic volume is known to vary from one subsection to the next. This obviously also represents a loss of information and a loss of statistical power to identify sources of variation in the number of accidents.

The choice of elementary unit of analysis must therefore be a compromise. In the three systems for safety analysis of road networks presented below, different choices have been made. In Germany, it is advised to use as long road sections as possible – limits of what is considered as possible being, for example, major changes in road layout, speed limit or traffic volume. In the United States, on the other hand, elementary road sections as short as 0.1 mile (0.16 km) are used, but a procedure for aggregating these into longer sections for the purpose of accident analysis has been developed. In Norway 1 kilometre road sections have been used.

In addition to the length of the road sections, the duration of the period to which data apply also influences the prospects for a successful analysis. A period of 3-5 years is commonly recommended, but in Norway a period of 8 years was used in the first network safety analysis that was performed.

4.1.2 Determining the treatment of classificatory variables in analysis

Roads are usually classified by administrative class and by function. For the purposes of network safety management, it is the functional classification that is of greatest interest. A choice to be made is whether to treat roads with different functional classifications as separate categories in network safety management, or merge all roads into the same system.

The use of accident prediction models is an important element of modern network safety management systems. If all types of road are included in the same accident prediction model, as has been done in Norway, variables should be included in that model to help identify the various types of road.

If, on the other hand, motorways, rural main roads, urban main roads, rural secondary roads and urban secondary roads are treated as separate categories, both the number of sections and the number of accidents will be reduced. Denmark has opted for an extensively classification of the road system, and the fitting of rather simple accident prediction models for each of the categories.

Any accident prediction model will, to certain extent, gloss over some details and will not identify all sources of systematic variation in the number of accidents. If a certain type of road has a very different safety record from the rest of the road system, it might therefore be wise to treat it as a category of its own, as it may be the case that an accident prediction model will not fully account for the special characteristics of the type of road.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

4.1.3 Developing a criterion for safety performance

The safety performance of a road can be described in terms of a number of different indices. Some candidates include:

1. The total expected number of accidents
2. The cost-weighted total expected number of accidents
3. The excess expected number of accidents for a certain section compared to an otherwise identical section with a normal expected number of accidents frequency
4. The cost-weighted excess number of accidents for a certain section compared to an otherwise identical section with a normal expected number of accidents and costs close to the mean cost of accidents
5. Prospective cost-effectiveness, which refers to the possibility of finding cost effective treatment for a road section
6. An abnormally high proportion of a specific type of accident or specific accident types.

It is not always necessary to make a choice between these criteria. Thus, one might initially identify roads that have a high total expected number of accidents (criterion 1). Next these roads might be ranked according to the cost of accidents (criterion 2). Analyses of potential safety treatments could be made for sections that had the highest cost-weighted expected number of accidents, and then these sections might be ranked by prospective cost-effectiveness (criterion 5). The various criteria can be applied at different stages of the network safety management process, permitting an effective use of all available information.

It is important to stress the fact that all criteria, except for the last one, are based on the expected number of accidents, not the recorded number of accidents. For large road systems, or long road sections, model-based estimates of the expected number of accidents may converge with the recorded number of accidents. For any given road section, on the other hand, this may not be the case. Whenever the recorded number of accidents is low (for example less than about 20), it provides a highly uncertain estimate of the expected number of accidents.

Confidence in the predictions for the expected number of accidents derived from accident prediction models has to depend, however, on whether these predictions are better than predictions based on using the recorded number of accidents. In a later section of this chapter, data from a study made in Norway will be presented to shed light on the question of how accurate estimates of the expected number of accidents based on the EB-method are.

Criteria 3 and 4 are quite similar to the usual criterion of a road accident black spot. These criteria appear to be theoretically attractive, since they recognise the facts that: (1) No accident prediction model will ever be able to perfectly predict the number of accidents for a given road section, since not all explanatory variables will be included in an accident prediction model, and (2) An excessive number of accidents at a given road section may be attributable to local risk factors present at that road section.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

However, the comparative study in terms of epidemiological criteria (sensitivity and specificity in black spot identification) in chapter 2 did not fully support this theory. On the contrary, it indicated that the risk factors not included in the accident prediction model do not always have stable effects over time on the expected number of accidents.

4.1.4 Identifying road sections with substandard safety performance

If an accident prediction model has been used, sections with substandard road safety performance can be identified by using the EB-method, as shown in chapter 2, dealing with black spot management. This method can of course be supplemented by statistical tests to make sure that an abnormally high expected number of accidents is not merely the result of chance variation.

In Norway, another approach was chosen. National roads were divided into three groups based on expected injury severity density (see below), hazardous road sections being defined as those that constituted the upper 10% of the distribution.

Irrespective of the approach chosen, road sections identified as hazardous are likely to form a distinct minority of all road sections that are included in the network safety management system.

4.1.5 Approach to the analysis of road sections with substandard safety performance

While hazardous road sections ought to be identified in terms of the expected number of accidents, accident analysis must be based on the recorded number of accidents. This introduces an element of inconsistency in network safety management. Whereas accident modelling and the identification of hazardous road locations, employing state-of-the-art techniques, fully recognises the major contribution that randomness makes to accidents, the traditional approach to accident analysis treats accident data as deterministic.

Suppose, for example, that the EB-estimate of the expected number of accidents for a road section is 9.6, but that 13 accidents have been recorded at this section. According to the logic of the EB-method, the 3.4 accidents that exceed the EB-estimate are the result of chance variation. How should we treat these accidents in accident analysis?

The traditional approach to accident analysis is fundamentally flawed and lacks credibility since it fails to recognise the contribution of randomness to accidents. There is, accordingly, a significant risk that it produces spurious explanations only. Chapter 2 indicated ways in which accident analysis at black spots can be improved to make them more rigorous and more open to empirical testing. Can something similar be applied to accident analysis as part of network safety management?

In Norway, the logic of the EB-approach to road safety analysis has been carried into the stage of accident analysis (Ragnøy and Elvik 2003). Analysis is initially performed in terms of recorded injury severity, i.e. the recorded number of injured road users weighted by the cost of the injuries. The objective of analysis is to identify the accident types that make the largest contribution to injury severity

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

density. Following an analysis in terms the recorded number of injured road users, results are adjusted by the EB-method so as to reflect long term values. An example of how this adjustment might function is given in Table 19.

Table 19: Adjusting recorded injury severity density to remove random variation. Source: Ragnøy and Elvik 2003

	Annual mean values based on 8 years of data				
Accident type	**Number of killed**	**Number of critically injured**	**Number of seriously injured**	**Number of slightly injured**	**Injury severity density**
Recorded values and recorded injury severity density					
Off the road				1.000	1.000
Rear-end				0.625	0.625
Head-on	0.125		0.250	1.625	7.665
Other				0.375	0.375
Total	0.125		0.250	3.625	9.665
Expected long term values for injury severity density					
Off the road		0.004		0.918	1.018
Rear-end		0.003		0.574	0.631
Head-on	0.083	0.008	0.186	1.492	5.831
Other		0.002		0.344	0.379
Total	0.083	0.016	0.186	3.328	7.852

Kilde: TØI report 883/2007

The adjusted values have removed the effects of regression-to-the-mean and therefore serve as the basis for estimating the effects of potential countermeasures.

4.2 A review of some systems for safety analysis of road networks

4.2.1 Network safety management in Germany

The following description of road network safety management in Germany is based on guidelines published by the Bundesanstalt für Strassenwesen and the French motorway directorate (Sétra) (BASt and Setra 2005). The objectives of safety analysis of road networks are:

1. to determine sections within the road network with a poor safety performance based on accident data and where deficits in road infrastructure have to be suspected and
2. to rank the sections by potential savings in accident costs in order to provide a priority list of sections to be treated by road administrations.

When these tasks have been performed, the following tasks are to analyse the accident structure of the sections in order to detect abnormal accident patterns

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

which can lead to possible improvement measures, and finally to offer the possibility to compare the costs of improvement measures to the potential savings in accident costs in order to rank measures by their benefit-cost ratio.

A distinction is made between three levels of accident severity:

1. Serious injury accidents, which includes fatal accidents
2. Slight injury accidents
3. Property-damage-only accidents.

The number of fatal accidents for a given road section is generally regarded as too low to provide a reliable basis for analysis. Hence, fatal accidents are considered in conjunction with accidents resulting in serious injury.

According to the German guidelines, road sections used in network screening should be as long as possible. However, a basic requirement is that each section must then be characterised by more or less the same traffic volume, the same cross section and the same type of environment (cross town link or rural section). It is recommended that the sections should be around 10 km (at least 3 km) long.

Four indicators of safety performance are used to help identify hazardous road sections:

1. Accident density, which is the number of accidents per kilometre of road per year
2. Accident cost density, which is the total societal cost of accidents per kilometre of road per year
3. Accident rate, which is the number of accidents per million vehicle kilometres of travel
4. Accident cost rate, which is the societal cost of accidents per million vehicle kilometres of travel.

All these indicators are stated in terms of the recorded number of accidents during a period of 3-5 years. No attempt is made to adjust for possible random fluctuations.

To identify road sections with poor safety performance, accident cost density is used. As resources are limited, those sections where improvements can be expected to have the highest benefit-cost ratio should be treated first. Therefore, information is needed on the accident costs per kilometre (or at a given location) and the safety potentials for possible remedial measures.

The safety potential (SAPO) is defined as the amount of accident costs per kilometre road length (cost density) that could be reduced if a road section had a best practice design. The higher the safety potential the more societal benefits can be expected from improvements of the road. The safety potential SAPO is calculated as the difference between the current accident cost density of the section ACD within the period under review and the basic accident cost density *bACD*.

The basic accident cost density *bACD* represents the anticipated average annual number and severity of road accidents (represented by the accident costs) per kilometre which can be achieved by a best practice design at the given average

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

daily traffic AADT. It is estimated as the product of basic accident cost rate *bACR* and average daily traffic AADT.

To make sure that the road sections identified as hazardous are not merely the result of random variation in accident counts, statistical tests are performed. The test consists of the comparison of the observed number of accidents A with the expected number of accidents *eA* of that section and the determination of the importance of the deviation by calculating the confidence interval of the observed values (Poisson law).

The sections of the road network are ranked according to the magnitude of the safety potential. Sections with a high rank are selected for more detailed engineering study designed to propose safety measures. It is recommended to present the results of the analysis in diagrams.

The detailed engineering study consists of an accident analysis. The initial stages of this analysis are identical to the analysis of road accident black spots. During the initial analyses, sites visits are not performed, but relevant information is extracted from computerised accident records.

4.2.2 Network safety management in Norway

The basic elements of the current Norwegian approach to safety analysis of road networks were developed in 2002 (Ragnøy, Christensen and Elvik 2002).

National roads were divided into 1-kilometre sections. The main question addressed in the new approach to the identification of hazardous road locations is this: How can road sections that have an abnormally high expected proportion of accidents resulting in fatal or serious injury be reliably identified?

In order to answer this question, the concept of injury severity density (ISD) was developed. The general definition of injury severity is given by this formula (Ragnøy, Christensen and Elvik 2002):

$$\text{ISD} = \frac{33.20FAT + 22.74CRI + 7.56SER + 1.00SLI}{Km \cdot year}$$

FAT = fatally injured road users (death within 30 days of accident)
CRI = critically injured road users
SER = seriously injured road users
SLI = slightly injured road users

These are the levels of injury severity used in official Norwegian road accident statistics. A critical injury is defined as an injury that is life-threatening or that leads to permanent impairment. A serious injury is one that is not critical, but generally requires treatment in hospital as an in-patient. Slight injuries are all those that are attended to by medical professionals, but will normally not require an overnight stay in hospital. For the country as a whole, about 2-3% of all injured road users recorded in official statistics are fatally injured, about 1-2% are critically injured, about 8-12% are seriously injured, and about 85-90% are

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

slightly injured. The total number of injured road users recorded per year is about 12,000 per year. In the above definition, km denotes kilometres of road and year denotes the number of years for which accident statistics are used in estimating injury severity density.

The weights assigned to each level of injury severity (33.20; 22.74, etc) are proportional to the societal costs of one injury of the stated severity (Elvik 1993). The cost of a fatal injury, or more precisely the value to society of preventing one, is roughly 33 times greater than the value of preventing a slight injury. When estimating injury severity density, one fatal injury is thus given 33.2 times the weight of one slight injury. In this way, fatal and serious injuries count for more than their numbers alone would imply.

In order to develop unbiased estimates of injury severity density for any road section, the empirical Bayes method was applied. In order to apply the empirical Bayes approach, multivariate models were fitted to explain the number of injured road users. Coefficient estimates for these models will be presented later. The empirical Bayes approach combines the estimates of the number of injured road users based on such models with the count of injured road users for a specific road section according to this equation:

$$\boldsymbol{E}(\boldsymbol{\lambda_i} / \boldsymbol{R_i}) = \boldsymbol{V_i} \cdot \boldsymbol{E}(\boldsymbol{\lambda_i}) + (1 - \boldsymbol{V_i}) \cdot \boldsymbol{R_i} \tag{23}$$

$E(\lambda|R_i)$ denotes the expected number of injured road users at a given level of injury severity (i = fatal, very serious , serious or slight), given that R_i injured road users were recorded. $E(\lambda_i)$ is the number of injured road users predicted by a multivariate model. V_i is the weight given to the predicted number of injured road users, $1 – V_i$ is the weight given to the recorded number of injured road users. According to Hauer (1997), the weight given to the predicted number of injured road users is given by:

$$\boldsymbol{V_i} = \frac{1}{1 + \dfrac{\boldsymbol{Var}(\boldsymbol{\lambda_i})}{\boldsymbol{E}(\boldsymbol{\lambda_i})}} \tag{24}$$

$Var(\lambda_i)$ is the systematic variation of the number of injured road users, $E(\lambda_i)$ is the predicted number. An estimate of $Var(\lambda_i)$ was obtained from the negative binomial regression models presented below.

The dependent variable in the models fitted was the number of injured road users of a given injury severity for each kilometre of road during a period of eight years. As noted above, injured road users in Norway are classified in four groups with respect to injury severity. Separate models were fitted for each level of injury severity. The models fitted had the following general form:

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

$$\lambda_i = \alpha \cdot AADT^b \cdot e^{\sum c \cdot x} \quad (25)$$

λ_i is the estimated number of injured road users of severity i (i = fatal, critical, serious or slight). α is a constant term, AADT is traffic volume raised to a power b, and e is the exponential function, that is the base of the natural logarithms (e = 2.71828), raised to a sum of parameters (c) for each of the explanatory variables (x). Models of a similar structure are widely used (for a recent example, see Taylor, Baruya and Kennedy 2002). The following explanatory variables were used (Ragnøy, Christensen and Elvik 2002):

1. Annual average daily traffic (AADT; a continuous variable)
2. Speed limit (50, 60, 70, 80 or 90 km/h)
3. The type of road, for roads that have a speed limit of 90 km/h (motorway class A, motorway class B, other road)
4. Number of lanes (1, 2, 3, etc)
5. Number of junctions per kilometre (0, 1, 2, etc)
6. Whether the road has the status of a national main road or not (yes/no)

Speed limit was represented by a set of dummy variables, using the speed limit of 50 km/h as the reference category. Whether or not a road was designated as a main road was also represented by a dummy variable. The other variables were entered in logarithmic form. For the number of lanes and the number of junctions per kilometre, 1 was added to the observed value to avoid taking the logarithm of zero. This was not done for AADT.

Analysis was based on data for 1-kilometre road sections of national roads in Norway for the period 1993-2000. The unit of analysis was 1 kilometre of road with data on accidents and explanatory variables for eight years (total for all eight years; not year-by-year). The length of national roads in Norway is about 26,500 kilometres. In the analyses, only sections with complete data on all explanatory variables were used. Moreover, these data had to take on constant values for the whole length of each 1-kilometre section. These restrictions resulted in a data set of 21,044 1-kilometre sections.

Table 20 shows the parameters of the models fitted, their standard errors, the exact P-value and statistics describing the explanatory power of each model.

It is seen that the expected number of killed or injured road users is closely related to traffic volume. The coefficient for traffic volume is quite stable, varying from 0.809 for seriously injured road users to 0.972 for slightly injured road users. The coefficients for speed limits fluctuate in both sign and magnitude and are not statistically significant for fatal and critical injuries. For serious and slight injuries, the coefficients indicate that the expected number of injured road users drops as speed limit goes up. It is highly unlikely that these coefficients show the true effects of raising speed limits; rather they reflect the fact the high speed limits are found on high-standard roads (see the instructive discussion of Taylor, Baruya and Kennedy 2002).

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Expected injury severity density was estimated for each of the 21,044 1-kilometre sections of road by combining model predictions according to equation 5 with the recorded number of injured road users for each road section.

Hazardous road sections are identified according to expected injury severity density. The Public Roads Administration of Norway has divided national roads into three classes according to expected injury severity density: (1) “Green roads”. These are the safest roads, that is those 50% of all national roads that have the lowest values for expected injury severity density and where no accidents resulting in fatalities or serious injuries have been recorded during the last eight years. (2) “Red roads”. The are the most hazardous roads, that is those 10% of all national roads that have the highest values for expected injury severity density, and where accidents resulting in fatalities or serious injuries have been recorded during the last eight years. (3) “Yellow roads”. These are the remaining 40% of the national roads, that are neither red nor green.

The analysis found that a higher percentage of main roads are red than of other national roads. 17.5% of main roads are red, versus 7.4% of other national roads. There is a very strong relationship between traffic volume and the probability that a main road is red. Mean AADT (annual average daily traffic) for red main roads is about 10,600, as opposed to only 600 for green main roads. Main roads that have an AADT of about 10,000 or more are often red throughout their whole length.

Once hazardous road sections were identified, an analysis of accidents is performed, using routinely available data only and not visiting each section. The objective of the accident analysis is to identify those accidents that make the greatest contribution to injury severity density.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 20: Parameter estimates for negative binomial accident prediction model for Norway

	Number of fatalities			Number of critically injured			Number of seriously injured			Number of slightly injured		
Variable	**Coefficient**	**Standard error**	**P-value**	**Coefficient**	**Standard error**	**P-value**	**Coefficient**	**Standard error**	**P-value**	**Coefficient**	**Standard error**	**P-value**
Constant	-7.154	0.530	0.000	-8.594	0.577	0.000	-6.778	0.301	0.000	-6.281	0.213	0.000
Ln (AADT)	0.842	0.036	0.000	0.829	0.047	0.000	0.809	0.021	0.000	0.972	0.012	0.000
Speed limit 60	-0.020	0.175	0.910	0.052	0.193	0.788	-0.393	0.090	0.000	-0.451	0.055	0.000
Speed limit 70	0.385	0.204	0.059	-0.009	0.244	0.969	-0.338	0.104	0.011	-0.311	0.066	0.000
Speed limit 80	0.172	0.165	0.299	0.161	0.180	0.369	-0.438	0.083	0.000	-0.506	0.049	0.000
Speed limit 90	0.090	0.222	0.686	0.025	0.263	0.923	-0.850	0.135	0.000	-0.743	0.069	0.000
Motorway class B	0.610	0.221	0.006	0.183	0.295	0.537	-0.466	0.148	0.002	-0.987	0.126	0.000
Motorway class A	0.879	0.775	0.256	-0.826	1.268	0.515	-1.155	0.430	0.007	-1.233	0.736	0.094
Ln (number of lanes + 1)	-1.967	0.449	0.000	-1.194	0.520	0.022	-0.523	0.272	0.055	-0.273	0.189	0.148
Ln (junctions/km + 1)	0.082	0.075	0.275	0.170	0.095	0.072	0.124	0.045	0.006	0.232	0.026	0.000
Main road dummy	0.255	0.069	0.000	0.245	0.096	0.011	0.047	0.043	0.270	-0.046	0.024	0.054
Proportion of random variation	0.662			0.808			0.502			0.076		
Proportion of systematic variation explained	0.236			0.132			0.364			0.804		

Kilde: TØI report 883/2007

Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

As part of the accident analysis, accidents were described according to these characteristics (Ragnøy and Elvik 2003):

1. Type of accident. A detailed numerical code was applied, identifying more than 50 different types of accident.
2. Road surface condition (bare dry, bare wet, snow or ice, etc.)
3. Weather conditions (sunny, overcast, raining, etc.)
4. Light conditions (daylight, dusk or dawn, darkness)

The analysis of accidents was based on the recorded number of accidents, or, more precisely, on the number of injured road users, specified according to injury severity in each type of accident. It was found that head-on crashes (frontal impacts) contribute substantially to overall injury severity density at those road sections that have the highest values for recorded injury severity density. Accordingly, safety treatments that can prevent or reduce the severity of frontal impacts are likely to be most effective in reducing injury severity density.

For some road sections, there was no type of accident making a dominant contribution to injury severity density; rather a disorderly pattern of accidents was observed, in which each type of accident made only a minor contribution to the overall score for injury severity density. For these road sections, the choice of effective safety treatments is more difficult and may require more in-depth studies.

4.2.3 SafetyAnalyst in the United States

SafetyAnalyst is a comprehensive software system for network safety management, developed by the United States Federal Highway Administration. Safety Analyst is described in four comprehensive white papers, comprising a few hundred pages. It is beyond the scope of this report to reproduce this detailed description. We refer to the four white papers, all of which can be downloaded from: www.safetyanalyst.org. The white papers are:

Module 1. Network screening (Harwood, Torbic, Bauer, Persaud, Lyon and Hauer 2002)
Module 2. Diagnosis and countermeasure selection (Harwood, Potts, Smiley, Bahar and Hauer 2002)
Module 3. Economic appraisal and priority-ranking (Harwood and Rabbani 2002)
Module 4. Evaluation (Harwood, Bauer and Torbic 2002)

Some key points from each module will be mentioned.

The objective of the network screening module is to use available data to review the entire roadway network under the jurisdiction of a particular highway agency and identify and prioritise those sites that need safety improvement. The network screening process relies on information on roadway characteristics and safety performance to identify those sites that are the strongest candidates for further investigation. The following types of data are used:

- Geometric design features
- Traffic control features
- Traffic volumes

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

- Accident history
- Accident characteristics
- Safety performance functions (SPFs)

To perform network screening, the first stage is to develop EB-estimates of the expected number of accidents by type and severity for each basic roadway element. A road section consists of multiple subsections or segments of varying length. The homogeneous section approach treats each of the segments, whatever its length, independently and assesses whether the safety performance of that segment is of sufficient concern to be selected for detailed engineering studies. The homogeneous section approach considers the safety performance of fixed-length segments within the homogeneous section and also includes a "peak searching" algorithm to identify the segments with highest accident frequency within a homogeneous section.

The peak searching algorithm is an innovative feature of Safety Analyst. It can be explained by means of an example. The implementation of this procedure requires a database where accidents can be allocated to subsections within each road section. With such a database, the road section is divided into 0.1-mile (0.16 kilometres) basic subsections as shown in Figure 12. A window consisting of W consecutive basic subsections is said to be of size W. Initially, the left edge of the window is placed at the left boundary of the road section, and the average expected accident frequency within the window is computed. The window is then moved one basic subsection to the right, and the average expected accident frequency is computed again. This is done until the right edge of the window reaches the right boundary of the road section. The process is repeated for windows of all feasible sizes. The largest of the averages so computed is the largest peak for a window of size W. Segment AA' in the figure is the highest peak when W = 3. Segment BB' is the second highest peak when W = 3. Segment CC' is the highest peak when W = 7. If two segments of size W overlap, only that with the higher estimated average is retained for further consideration. For example, if on segment AA' the statistical precision criterion is met, segment AA' rather than segment CC' will be considered further. The reason is that the rank of segment AA' is bound to be higher than the rank of CC' and, if a diagnosis is conducted, the vicinity of AA' will be considered when a project is formulated. If the statistical precision criteria on AA' are not met but they are met on CC', only the latter will be retained for ranking. To specify the required statistical precision, use of "limiting coefficients of variation" is suggested. The coefficient of variation of an estimate, denoted as "*CV*," is given by:

CV = (standard error of estimate) / (expected estimate)

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

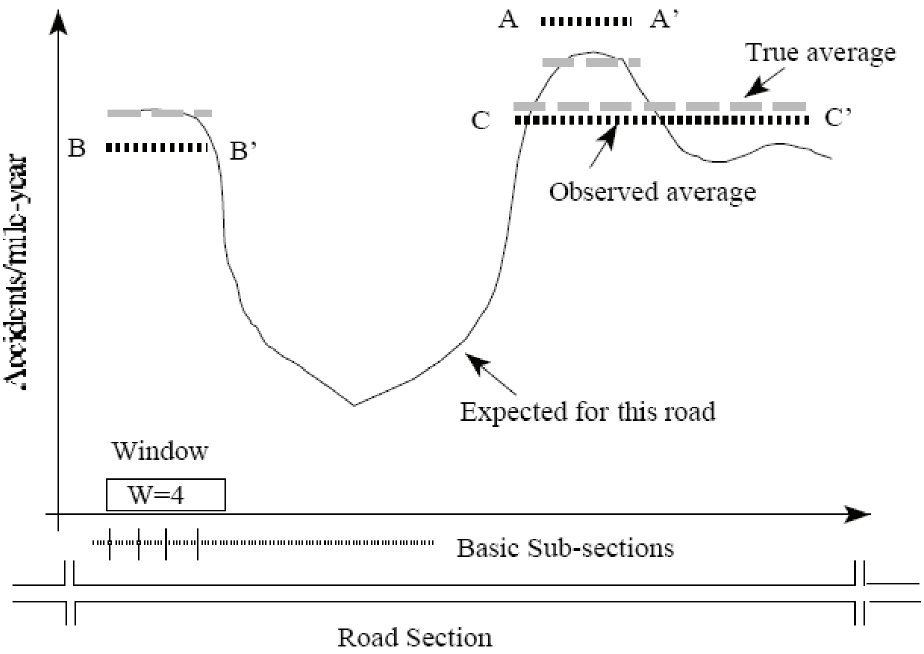
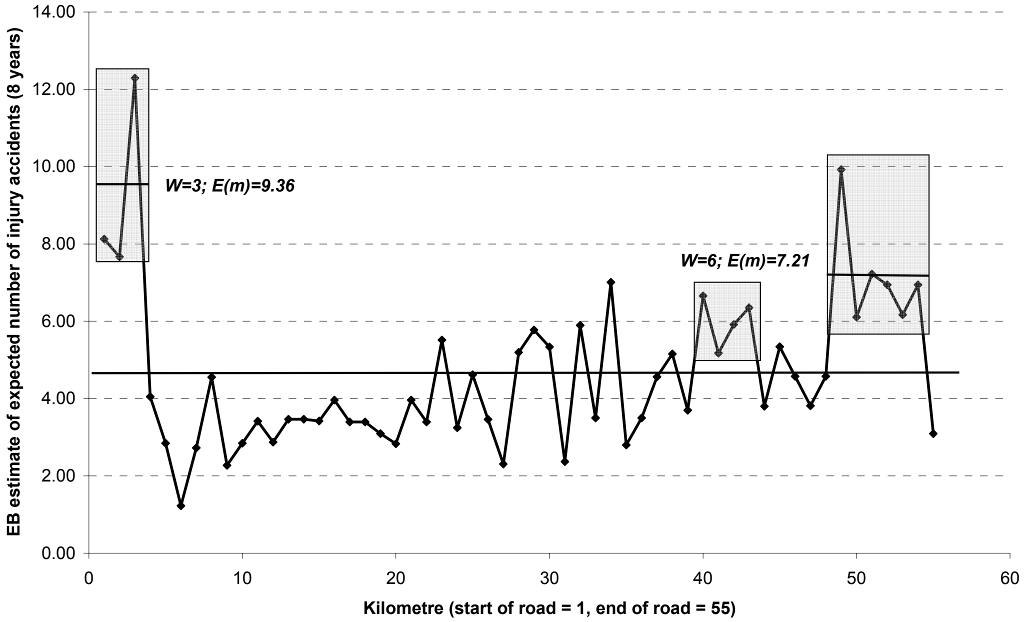


Figure 12: Peak searching algorithm in Safety Analyst. Source: Safety Analyst website.

Once peaks have been identified, sites are ranked for detailed study by prospective cost-effectiveness of potential safety improvements. The prospective effectiveness can be based on either the expected accident frequency or the expected excess accident frequency. Prospective cost-effectiveness can be based on average costs of particular project types and benefits can be assumed to be proportional to either expected accident frequency or expected excess accident frequency.

The profiles-and-peaks routine can be illustrated by means of an example. Figure 13 shows EB-estimates for the number of injury accidents for a 55 kilometre long section of national road in Norway. The example is based on the description of the profiles and peaks method given by Hauer (2000).

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961



Kilde: TØI report 883/2007

Figure 13: EB-estimates of the expected number of accidents per kilometre for a 55 kilometre long national road in Norway

As can be seen, the expected number of accidents varies considerably along the road. The mean EB-expected number of accidents for the entire road was 4.68 per kilometre. In Figure 13, a few sections that appear to have an abnormally high expected number of accidents have been marked by shaded rectangles. The question is: can these sections be treated as single peaks and accidents analysed for the entire shaded sections as a whole?

To answer this question, the standard error of the EB-estimates was estimated for each kilometre of road. Then windows of size 3, 4 and 6 were used to identify potential peaks. A window of size 3 was judged appropriate for the section to the left of the figure, a window of size 4 was judged appropriate for the section starting at kilometre 40. Finally a window of size 6 was judged appropriate for the section starting at kilometre 49 (to the right in Figure 13). The critical value for the coefficient of variation was set to 0.10.

By applying the windows, the sections to the left and right of Figure 13 passed the criterion; the section around kilometre 40 did not pass the criterion. For the section to the left, the total recorded number of accidents was 31, which was considerably higher than the normal number (model estimate) of 21.8. The EB-estimate was 28.1. For the section to the right, the total recorded number of accidents was 38; the normal number was 62 and the EB-estimate was 43.3. Thus, this section did not have an abnormally high number of accidents compared to what was expected for it. It only appeared to have an abnormally high number of accidents when compared to the mean EB-estimate for the entire road.

Module 2 of Safety Analyst, diagnosis and countermeasure selection, points out that: *"The nature of accidents is that they are rare, multi-causal, and random. Because accidents are rare, it follows that the presence of accident patterns*

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

provides compelling evidence of underlying safety deficiencies. Because accidents are random, it also follows that the accident history at a given site will only provide partial information about safety at that site. Experience tells us that:

- *Safety deficiencies can contribute to serious accidents, for which there was no evidence by way of a previous pattern.*
- *Some sites with high numbers of accidents do not have readily identifiable accident patterns.*
- *There can be evidence in the accident history that a given deficiency has contributed to accidents at one site while at another site, with a similar deficiency, there is no clear pattern of associated accidents.*
- *A given deficiency can contribute to different accident types."*

For each specific location to be investigated, Safety Analyst will perform the following sequence of steps:

1. Prepare collision diagram template
2. Plot collision diagram
3. Identify accident patterns
4. Diagnose safety problems
5. Identify and select appropriate countermeasures

To help diagnose factors contributing to accidents, Safety Analyst asks a number of diagnostic questions. An example of a diagnostic question for single-vehicle accidents on a horizontal curve is as follows:

> *When drivers reach a curve, which is much sharper than curves on the preceding road section, especially if they are unfamiliar, they can be surprised and find themselves approaching it at too high a speed. Is the curvature of the accident site unusually sharp in relation to the previous several miles of road section?*

An investigator familiar with the road network may be able to answer this question on the basis of his or her own knowledge, or by using a photolog or accessing design drawings. If the answer to this question is no, then it will not be flagged as a diagnosis for further investigation. Otherwise, it will be flagged and will be included in items to investigate further during the site visit.

Each diagnostic question is framed to lead to a diagnosis and each diagnosis is framed to lead to a countermeasure. The diagnostic questions are intended to cover typical accident scenarios, rather than rare and unusual situations.

The output of the office investigation diagnosis stage will be in the form of an interim report comprising all assembled information, such as:

- Collision diagram with all characteristics (traffic devices, geometrics, etc.) to be taken to the site if that investigation will take place
- List of accident patterns to be diagnosed, annotated with the season and time of the week and of the day, if any, that the accidents are most prevalent

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

- List of questions about the site that could not be answered in the office
- List of initial diagnoses
- List of potential countermeasures to be considered in the field

Following a field visit to each site, a final proposal for countermeasures is developed. Countermeasure selection involves multiple technical and budgetary considerations that are not appropriate for automated decision making.

For each concern identified through the diagnostic process, a list of potential countermeasures will be selected. If more than one concern has been identified at the site, more than one countermeasure list will be generated.

4.3 The performance of the Empirical Bayes method

In Norway and the United States, network safety management relies on the empirical Bayes (EB) method. In Germany this method is not used.

There are a number of problems in using the EB-method, in particular when using it to evaluate the effects of safety measures (Persaud and Lyon 2007). Here, we would like to know how well it performs when predicting the number of accidents expected to occur without treatment.

In this section, the predictive accuracy of EB-estimates of the expected number of accidents derived in four ways is compared:

1. Estimates derived from the empirical distribution of accidents in a population of sites, employing the method of moments.
2. Estimates derived from the parameters of a negative binomial distribution fitted to an empirical distribution of accidents in a population of sites by means of the maximum likelihood technique.
3. Estimates derived by combining the predictions of a simple accident prediction model and the recorded number of accidents for a site.
4. Estimates derived by combining the predictions of a more elaborate accident prediction model and the recorded number of accidents for a site.

The four versions represent increasing levels of statistical refinement and complexity. All four versions of EB-estimates are compared to the traditional, naïve assumption of treating the recorded number of accidents as an unbiased estimate of the expected number of accidents.

4.3.1 Data and methods

Two data sets are used to test the predictive accuracy of EB-estimates. Short descriptions of these are given below.

4.3.1.1 Norwegian data

The Norwegian data refer to 21,738 1-kilometre sections on national roads. For these road sections, data on accidents and a number of variables associated with the number of accidents were obtained for the period 1997-2004. This period was

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

divided into two periods: 1997-2000 and 2001-2004. All road sections existed throughout both periods. The road sections remained unchanged, except for ordinary road maintenance, like resurfacing, renewing road markings and replacing traffic signs. On a few road sections, more extensive treatments may have been introduced. It is, however, reasonable to assume that more extensive treatments were applied only to a few road sections (less than 10 %), and that such treatments took place at a nearly constant rate throughout the entire period (Elvik and Rydningen 2002). We may thus treat this population of road sections as suitable for testing the accuracy of EB-estimates designed to answer the question: what would have happened if no measure had been introduced?

Table 21 shows the empirical distribution of road sections by the number of accidents during the period 1997-2000. The distribution has a very long tail, which is not shown in full. The maximum number of accidents recorded was 47. Negative binomial distributions have been fitted to the empirical distribution by means of the method of moments and the maximum likelihood method. The empirical distribution deviates significantly from both these negative binomial distributions. The mean number of accidents was 0.589, the variance was 2.547.

Two accident prediction models have been fitted to the data. The simplest of these models was of the form:

Simple model = $\alpha \cdot \mathrm{AADT}^{\beta} e^{\Sigma \gamma_i x_i}$

AADT denotes annual average daily traffic, e is the base of natural logarithms and α, β and γ are coefficients estimated by maximum likelihood techniques. The explanatory variables, in addition to traffic volume, included speed limit, type of road (freeway versus other), number of lanes, number of intersections per kilometre and a dummy for trunk road status. The more advanced model was of the form:

Advanced model = $\alpha \cdot \mathrm{AADT}^{\beta} e^{\varepsilon(\mathrm{AADT}/1000)} e^{\Sigma \gamma_i x_i}$

All explanatory variables were the same as in the simple model, but the advanced model allows for a different shape of the relationship between traffic volume and the number of accidents.

4.3.1.2 Portuguese data

The Portuguese data refer to 3,470 road sections on rural two-lane roads. Most of these had a length of 250 metres. For these road sections, data on accidents and other variables were available for the period 1994-2003. This period was divided into the two periods 1994-1998 and 1999-2003. Accident prediction models of the same form as shown above for Norway were fitted for the period 1994-1998 and estimates of the expected number of accidents for the period 1999-2003 obtained. Explanatory variables included AADT, road width and section length.

Table 23 shows the empirical distribution of sections by number of accidents during 1994-1998. Negative binomial distributions were fitted to the empirical distribution by means of the method of moments and the maximum likelihood method. Unlike the Norwegian data, the negative binomial distributions fit well to the empirical distribution in the Portuguese data.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

4.3.2 Results

4.3.2.1 Norwegian data

Table 22 shows results for the Norwegian data set. Due to the long tail of the distribution, the table does not show all results for road sections that had 15 or more accidents in the first period. In most cases, EB-estimates are more accurate than using the count of accidents in the first period as an estimate of the expected number of accidents in the second period. A total of 34 predictions were made, for the counts of accidents represented in the data set (all counts from 0 through 15 were represented and selected values between 16 and 47, making for 34 in total). EB-estimates were more accurate than the count of accidents in 29 of these 34 cases.

Estimates based on the simple accident prediction model were the most accurate in 11 cases; those based on the more advanced accident prediction model were the most accurate in 4 cases. EB-estimates not based on accident prediction models were the most accurate in 14 cases. Thus, predictive accuracy can be improved by using and accident prediction model, but this is not consistently the case. Moreover, an advanced accident prediction model does not necessarily improve predictions compared to a simpler accident prediction model.

Figure 14 shows the percentage prediction error. It can be seen that the error associated with EB-estimates fluctuates randomly around the correct values. There is no tendency for EB-estimates to consistently predict too low or too high numbers. Using the accident count for the first period, on the other hand, yields consistently too high estimates for all sites that recorded any accidents at all during the first period.

4.3.2.2 Portuguese data

Table 24 shows results for the Portuguese data. A total of 17 predictions were made for sections that recorded 0, 1, 2, …, 17 accidents in the first period. Predictions based on EB-estimates are more accurate than predictions based on the count of accidents in all cases. However, none of the four versions of EB-estimates provides consistently more accurate predictions than the others. Predictions based on maximum likelihood estimates of the parameters of the negative binomial distributions are the most accurate in 8 cases; model based predictions are the most accurate in 9 cases. An advanced accident prediction model is not associated with an improved predictive performance compared to a simple accident prediction model.

Figure 15 shows percentage prediction errors. It is seen the all the predictions are too high for sites that recorded more than 2 accidents in the first period, although the EB-estimates are closer to the actual values than the count of accidents in the first period. In other words, EB-estimates do not adjust fully for regression-to-the-mean in this data set. Relying on EB-estimates in a before-and-after study using this data set is therefore likely to have resulted in inflated estimates of the effects of safety measures.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

4.3.3 Discussion and conclusions

The empirical Bayes approach to road safety estimation has almost become the gold standard for observational before-and-after studies of road safety measures in the past 10 years. Before accepting the EB-approach as the gold standard we need to know how accurate it is. Does the approach really control adequately for regression-to-the-mean?

The study presented here shows that the EB-approach nearly always provides better predictions of the number of accidents than the traditional approach of assuming that the recorded number of accidents is an unbiased estimator of the expected number of accidents. EB-estimates are, however, not always accurate. If the differences between EB-estimates and the actual number of accidents are small and random, such inaccuracies are fully acceptable – indeed inevitable, given the randomness inherent in accident counts. The results based on Norwegian data suggest that EB-estimates are in most cases very accurate and that the errors are unsystematic. This is reassuring and supports use of the EB-approach.

The findings based on Portuguese data were somewhat less reassuring. Although predictions based on EB-estimates were clearly better than those based on the count of accidents, they nevertheless were systematically too high for road sections that recorded more than 2 accidents in the first period. In the Portuguese data, EB-estimates did not fully remove regression-to-the-mean. It is not clear why this was the case. These findings should prompt further research and do not by themselves constitute a sufficient reason for rejecting the EB-approach. The EB-approach should be used until a better approach has been developed. It should be noted that as accident prediction models become ever more complex and refined, they may lead to more precise predictions, in particular by allowing the over-dispersion parameter to vary. The comparison in this paper was confined to models with a fixed over-dispersion parameter, differing only with respect to the parameters describing the shape of the relationship between traffic volume and the number of accidents.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Table 21: Distribution of accidents during the period 1997-2000 on 1-kilometre road sections on national roads in Norway

Count of accidents 1994-1998	Number of sections	Negative binomial distribution, fitted by method of moments	Negative binomial distribution, fitted by maximum likelihood technique
0	15178	16818	15452
1	4110	2263	3344
2	1238	1023	1414
3	539	571	695
4	240	349	366
5	148	224	200
6	78	149	112
7	57	101	64
8	30	70	37
9	24	49	22
10	20	34	13
11	11	24	8
12	5	18	4
13	8	13	3
14	11	9	2
15-19	18	20	2
20-47	23	4	0
Total	21738	21737	21737
Mean number of accidents	0.5891	$X^2 = 1898.196$; df = 16; p = 0.000	$X^2 = 591.883$; df = 12; p = 0.000
Variance of number of accidents	2.5467		

Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

TABLE 22: Prediction of number of accidents during the period 2001-2004 on 1-kilometre road sections in Norway based on EB-estimates derived from data for the period 1997-2000

				Alternative predictions of the number of accidents during 2001-2004				
Count of accidents 1994-1998	**Number of sections**	**Total number of accidents 1997-2000**	**Total number of accidents 2001-2004**	**1997-2000 count of accidents**	**Method of moments empirical distribution**	**Maximum likeli-hood negative binomial parameters**	**Simple accident prediction model**	**Refined accident prediction model**
0	15178	0	4103	0	3235	2011	***3919***	3911
1	4110	4110	2559	4110	3423	3659	***2574***	2591
2	1238	2476	1547	2476	1798	2041	1461	***1470***
3	539	1617	971	1617	1117	1297	***1051***	1057
4	240	960	562	960	646	759	***642***	645
5	148	740	556	740	490	580	***555***	552
6	78	468	314	468	***307***	365	354	351
7	57	399	344	399	259	310	***325***	322
8	30	240	214	240	155	186	***200***	197
9	24	216	196	216	139	167	***183***	183
10	20	200	164	200	128	154	171	***167***
11	11	121	89	121	77	***93***	105	104
12	5	60	59	***60***	38	46	50	51
13	8	104	68	104	***66***	80	95	93
14	11	154	137	154	98	118	***134***	134
15-19	18	300	234	300	190	230	262	258
20-47	23	640	493	640	402	488	529	524
Best predictions				5	9	5	11	4

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

TABLE 23: Distribution of accidents during the period 1994-1998 on 250-metre road sections on rural two-lane roads in Portugal

Count of accidents 1994-1998	Number of sections	Negative binomial distribution, fitted by method of moments	Negative binomial distribution, fitted by maximum likelihood technique
0	2069	2058	2064
1	586	619	615
2	335	310	307
3	177	175	175
4	104	105	105
5	60	65	65
6	42	41	41
7	32	27	27
8	14	17	17
9	8	11	11
10	4	7	8
11	8	5	5
12	2	3	3
13	4	2	2
14	3	2	2
15	2	1	1
17	1	0	0
Total	3451	3449	3449
Mean number of accidents	1.0014	$X^2 = 11.937$; df = 11; p = 0.368	$X^2 = 11.726$; df = 12; p = 0.468
Variance of number of accidents	3.3338		

Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

TABLE 24: Prediction of number of accidents during the period 1999-2003 on 250-metre road sections in Portugal based on EB-estimates derived from data for the period 1994-1998

				Alternative predictions of the number of accidents during 1999-2003				
Count of accidents 1994-1998	**Number of sections**	**Total number of accidents 1994-1998**	**Total number of accidents 1999-2003**	**1994-1998 count of accidents**	**Method of moments empirical distribution**	**Maximum likeli-hood negative binomial parameters**	**Simple accident prediction model**	**Refined accident prediction model**
0	2069	0	1275	0	668	675	728	***728***
1	586	586	600	586	636	636	606	***604***
2	335	670	589	670	619	617	605	***602***
3	177	531	376	531	462	461	445	***442***
4	104	416	287	416	351	***349***	350	351
5	60	300	161	300	248	247	***239***	241
6	42	252	122	252	206	***205***	205	209
7	32	224	124	224	181	180	***178***	179
8	14	112	57	112	90	***90***	93	92
9	8	72	29	72	57	57	58	***57***
10	4	40	21	40	32	***32***	35	36
11	8	88	28	88	70	***69***	73	74
12	2	24	12	24	19	***19***	20	20
13	4	52	18	52	41	***41***	44	46
14	3	42	23	42	33	33	32	***32***
15	2	30	16	30	24	23	22	***19***
17	1	17	9	17	13	***13***	15	16
Best predictions				0	0	8	2	7

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

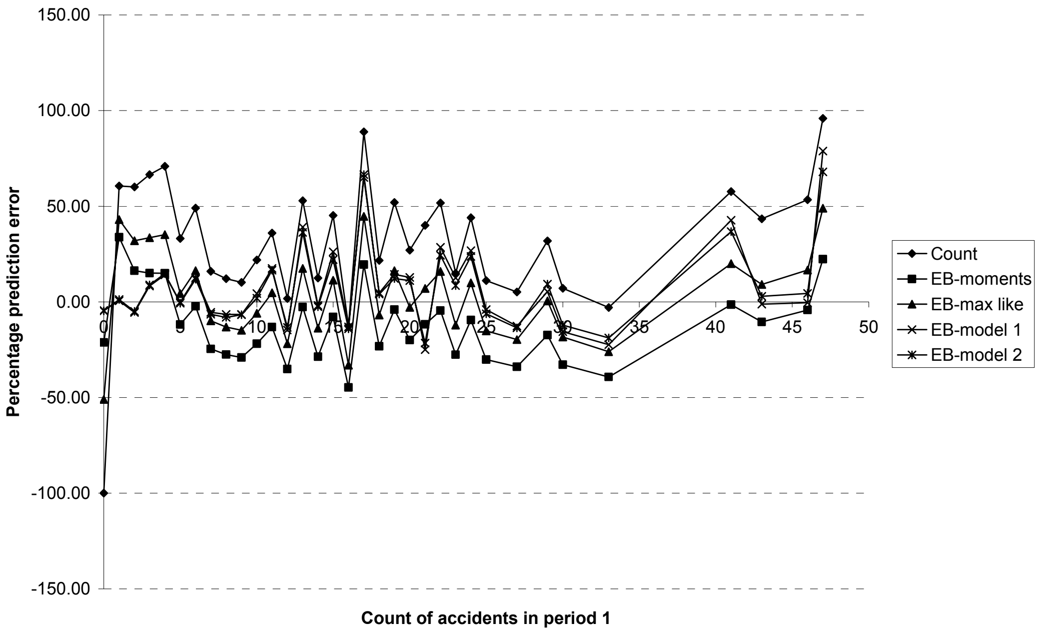


Figure 14: Prediction error for EB-estimates for Norway

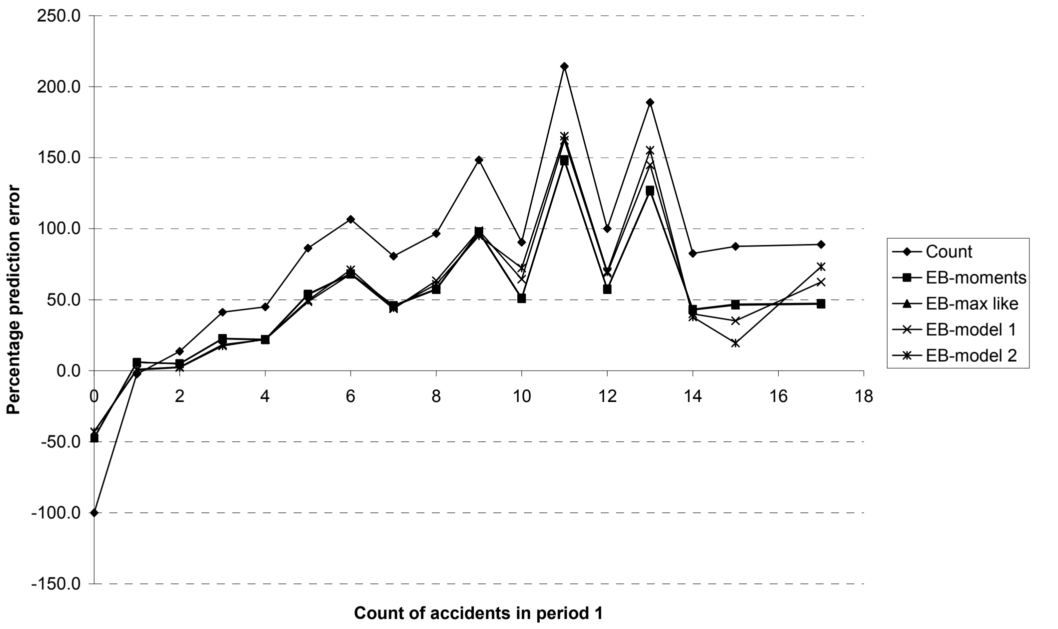


Figure 15: Prediction error for EB-estimates for Portugal

4.4 Summary of key elements of safety analysis of road networks

This chapter has reviewed key elements of safety analysis of road networks, which is a basic function of network safety management. The main objective of safety analysis of road networks is to identify road sections that have a potential for cost-effective improvements in road safety. Three systems for safety analysis

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

of road networks have been reviewed: The German system, the Norwegian system and the Safety Analyst system of the United States.

Neither of these systems correspond to state-of-the-art practice. However, all systems have important elements in common, including:

1. Safety analysis of road networks normally comprises an extensive network of roads of at least several hundred, but more commonly several thousand kilometres.
2. Roads are to a certain extent pre-classified by type, such as motorways, rural main roads and urban main roads.
3. For analysis, elementary units of analysis are defined. The elementary units should ideally speaking be completely homogeneous (within each unit) with respect to factors influencing the number of accidents. Whereas very short basic units are used in the United States, the approach in Germany is to make each section as long as possible.
4. Safety analysis of road networks is designed to account for as many sources of variation in the number of accidents as possible. In Norway and the United States, this is done by means of accident prediction models. In Germany, various measures of safety performance are estimated as rates using traffic volume as the denominator.
5. Road sections with a substandard safety performance can be identified in a number of ways. All three countries employ estimates of the costs of accidents to describe safety performance.
6. In the United States, a routine has been developed for merging adjacent short sections into longer sections to obtain a better basis for accident analyses (the profiles and peaks method). A similar routine has not been developed in Norway and Germany.
7. Analysis of accidents for sections identified as hazardous relies on the recorded number of accidents in all countries. In Norway, the results of analyses are adjusted to remove regression-to-the-mean. Similar adjustments are, as far as is known, not applied in Germany and the United States.
8. Analysis of accidents is in all countries done in two stages: a preliminary analysis in the office, followed by a more detailed analysis by means of site visits.

The system for safety analysis of road networks is very similar in Norway and the United States. The German system relying on accident rates or accident cost rates represents a different approach, not requiring the use of accident models. As discussed in detail in Chapter 3, accident modelling can be difficult and there are many pitfalls. Yet, on balance, the prevailing opinion today seems to be that approaches based on accident modelling are to be preferred, since they permit controlling for randomness in accident counts by means of the empirical Bayes method.

The approach to accident analysis needs considerable development. Current techniques, as discussed in chapter 2, involve a number of uncertainties and the

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

conclusions drawn on the basis of traditional accident analyses have never actually been scrutinised critically, but seem to have been accepted by almost everybody as obviously correct. As yet, however, no fully satisfactory approach to accident analysis for hazardous road locations has been developed.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

5 Discussion and conclusions

The objective of this report was to describe state-of-the-art approaches to black spot management and safety analysis of road networks. By this is meant approaches that represent the best conceivable practice with respect to black spot management and safety analysis of road networks.

Black spot management has a long tradition in traffic engineering. Yet, the approaches taken by different countries to black spot management are strikingly different. There is, accordingly, a need for discussing what constitutes state-of-the-art practice.

Research during the past 20 years has questioned many elements of the traditional approach to black spot management. Although a number of studies have cast doubt on some approaches to black spot management, this should be seen as a first input towards developing better approaches.

The essential elements of an emerging state-of-the-art are as follows:

1. Black spots should be identified in terms of the expected number of accidents, not the recorded number of accidents.
2. Black spots should be identified by reference to a clearly defined population of sites, whose members can in principle be enumerated.
3. Use of a sliding window approach to identifying black spots is discouraged. This approach artificially inflates variation in accident counts.
4. To estimate the expected number of accidents, multivariate accident prediction models should be developed.
5. The best estimate of the expected number of accidents for a single site is obtained by combining the recorded number of accidents with the model estimate for that site. This should be done by applying the empirical Bayes method.
6. The performance of alternative critical values for the expected number of accidents qualifying a site as black spot should be investigated in terms of sensitivity and specificity. An optimal criterion should be chosen.
7. The traditional criterion for a true black spot, which is that there is a dominant pattern of accidents, has not been validated. Analysis of accidents at black spots is best viewed as a means of developing hypotheses regarding potentially contributing factors to the accidents.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

8. Analysis of black spots should recognise the possibility that an apparent pattern may arise as a result of chance alone. Binomial tests should be applied to determine the probability that a certain number of accidents of a certain type is the result of chance only.
9. Analysis of black spots should employ a blinded design and rely on a comparison of the black spot to a safe location. The task of analysts is to identify risk factors for accidents. Analysts should not known which site is the black spot and which site is the safe one.
10. Evaluation of the effects of black spot treatment should employ the empirical Bayes before-and-after design.

A state-of-the-art approach to safety analysis of road networks should contain all these elements. In addition, a state-of-the-art approach to safety analysis of road networks should include a routine for merging adjacent sections for the purpose of accident analysis. The profiles and peaks algorithm is suitable for this purpose.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

6 References

Asher, H. B. (1976). *Causal modelling*. Sage Publications, Beverly Hills, CA.

Brüde, U. And J. Larsson (1982). *"Farliga" korsningar på det statliga vägnätet. Analys och åtgärdsförslag*. VTI-meddelande 299. Statens Väg- och Trafikinstitut (VTI), Linköping.

Brüde, U. and J. Larsson (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit? *Accident Analysis and Prevention*, **25**, 499-509.

Cresswell, W. L. and P. Froggatt (1963). *The causation of bus driver accidents. An epidemiological study*. Oxford University Press, London.

Danielsson, S. (1988). Estimation of the effects of countermeasures on different types of accidents in the presence of regression effects. *Accident Analysis and Prevention*, **20**, 289-298.

Deeks, J. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. Chapter 14, 248-282, in: Egger, M., G. Davey Smith and D. G. Altman (Eds): *Systematic reviews in Health care. Meta-analysis in context*. BMJ books, Second edition, London.

Elvik, R. (1988). Some difficulties in defining populations of "entities" for estimating the expected number of accidents. *Accident Analysis and Prevention*, **20**, 261-275.

Elvik, R. (1993). *Økonomisk verdsetting av velferdstap ved trafikkulykker. Dokumentasjonsrapport*. TØI-rapport 203. Transportøkonomisk institutt, Oslo.

Elvik, R. (1997). Evaluations of road accident black spot treatment: a case of the Iron Law of evaluation studies? *Accident Analysis and Prevention*, **29**, 191-199.

Elvik, R. (2001). *Quantified road safety targets. An assessment of evaluation methodology*. Report 539. Institute of Transport Economics, Oslo.

Elvik, R. (2004). Traffic safety. Chapter 16 in M. Kutz (Ed): *Handbook of Transportation Engineering*. McGrawHill, New York, NY.

Elvik, R. (2006). New approach to accident analysis for hazardous road locations. *Transportation Research Record*, **1953**, 50-55.

Elvik, R. (2007). Operational criteria of causality for observational road safety evaluation studies. Paper TRB-07-0291. Forthcoming in *Transportation Research Record*.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Elvik, R. and U. Rydningen (2002). *Effektkatalog for trafikksikkerhetstiltak*. Rapport 572. Transportøkonomisk institutt, Oslo.

Fridstrøm, L., J. Ifver, S. Ingebrigtsen, R. Kulmala and L. K. Thomsen (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis and Prevention*, **27**, 1-20.

Fridstrøm, L. (1999). *Econometric models of road use, accidents, and road investment decisions*. Volume II. Report 457. Institute of Transport Economics, Oslo.

Gaudry, M. and S. Lassarre (Eds) (2000). *Structural road accident models. The international DRAG family*. Pergamon Press, Oxford.

German road and transport research association (2006). *Recommendations on accident type maps including detailed definitions regarding black spot detection and analysis*. Cologne.

Geurts, K. (2006). *Ranking and profiling dangerous accident locations using data mining and statistical techniques*. Doctoral dissertation. Faculty of applied economics, Hasselt University, Hasselt.

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Analysis and Prevention*, **35**, 273-285.

Hakkert, A. S. and L. Braimaister (2002). *The uses of exposure and risk in road safety studies*. Report R-2002-12. SWOV Institute for Road Safety Research, Leidschendam.

Harwood, D. W., K. M. Bauer and D. J. Torbic (2002). *Safety Analyst: Software tools for safety management of specific highway sites. White paper for module 4 – evaluation*. Federal Highway Administration, Washington DC.

Harwood, D. W., I. B. Potts, A. Smiley, G. Bahar and E. Hauer (2002). *Safety Analyst: Software tools for safety management of specific highway sites. White paper for module 2 – diagnosis and countermeasure selection*. Federal Highway Administration, Washington DC.

Harwood, D. W. and E. K. Rabbani (2002). *Safety Analyst: Software tools for safety management of specific highway sites. White paper for module 3 – economic appraisal and priority ranking*. Federal Highway Administration, Washington DC.

Harwood, D. W., D. J. Torbic, K. M. Bauer, B. N. Persaud, C. A. Lyon, and E. Hauer (2002). *Safety Analyst: Software tools for safety management of specific highway sites. White paper for module 1 – network screening*. Federal Highway Administration, Washington DC.

Hauer, E. (1986). On the estimation of the expected number of accidents. *Accident Analysis and Prevention*, **18**, 1-12.

Hauer, E. (1995). On exposure and accident rate. *Traffic Engineering and Control*, **36**, 134-138.

Hauer, E. (1997). *Observational before-after studies in road safety.* Pergamon Press, Oxford.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Hauer, E. (2000). *Working paper 10. Screening: methods and software.* Unpublished paper prepared for task force developing Safety Analyst. Available at: www.roadsafetyresearch.com

Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis and Prevention*, **33**, 799-808.

Hauer, E. (2004). Statistical road safety modelling. *Transportation Research Record*, **1897**, 81-87.

Hauer, E. (2005A). *Cause and effect in observational cross-section studies on road safety*. Unpublished manuscript, dated March 1, 2005.

Hauer, E. (2005B). The road ahead. *Journal of Transportation Engineering*, **131**, 333-339.

Hauer, A., B. K. Allery, J. Kononov and M. S. Griffith (2004). How best to rank sites with promise. *Transportation Research Record*, **1897**, 48-54.

Hauer, E. and J. Bamfo (1997). Two tools for finding what function links the dependent variable to the explanatory variables. Paper presented at ICTCT workshop in Lund, Sweden, November 5-7, 1997. Available at: www.roadsafetyresearch.com

Hauer, E., F. Council and Y. Mohammedshah (2004). Safety models for urban four-lane undivided road segments. *Transportation Research Record*, **1897**, 96-105.

Hauer, E., D. W. Harwood, F. M. Council and M. S. Griffith (2002B). Estimating safety by the empirical Bayes method. A tutorial. *Transportation Research Record*, **1784**, 126-131.

Hauer, E., J. Kononov, B. Allery and M. S. Griffith (2002A). Screening the road network for sites with promise. *Transportation Research Record*, **1784**, 27-32.

Hauer, E. and B. N. Persaud (1984). Problem of identifying hazardous road locations using accident data. *Transportation Research Record*, **975**, 36-43.

Hauer and Quaye (1990). On the use of accident or conviction counts to trigger action. Unpublished manuscript. University of Toronto, Safety Studies Group, Department of Civil Engineering.

Hauer, E., K. Quaye and Z. Liu (1993). On the use of accident or conviction counts to trigger action. *Transportation Research Record*, **1401**, 17-25.

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. The Free Press, New York, NY.

Ivan, J. N. (2004). New approach for including traffic volumes in crash rate analysis and forecasting. *Transportation Research Record*, **1897**, 134-141.

Jarrett, D. F., C. R. Abbess and C. C. Wright (1988). *Empirical estimation of the regression-to-mean effect associated with road accident remedial treatment*. Proceedings of Traffic Safety Theory and Research Methods, Session 2, Models for Evaluation. SWOV Institute for Road Safety Research, Leidschendam.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Jonsson, T. (2005). Predictive models for accidents on urban links. Doctoral dissertation. Bulletin 226. Lund Institute of Technology, Department of Technology and Society, Traffic Engineering, Lund.

Kim, D-G. and S. Washington (2006). The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. *Accident Analysis and Prevention*, **38**, 1094-1100.

Kononov, J. (2002). Identifying locations with potential for accident reductions. Use of direct diagnostics and pattern recognition methodologies. *Transportation Research Record*, **1784**, 153-158.

Lee, J. and F. L. Mannering (2002). Impact of roadside features on the frequency and severity of run-off-road accidents: an empirical analysis. *Accident Analysis and Prevention*, **34**, 349-361.

Lemaire, J. (1995). *Bonus-malus systems in automobile insurance*. Kluwer Academic Publishers, Boston, MA.

Lenguerrand, E., J. L. Martin and B. Laumon (2006). Modelling the hierarchical structure of road accident data – application to severity analysis. *Accident Analysis and Prevention*, **38**, 43-53.

Lord, D. and B. N. Persaud (2000). Accident prediction models with and without trend. Application of the generalized estimating equations procedure. *Transportation Research Record*, **1717**, 102-108.

Lord, D.; S. P. Washington and J. N. Ivan (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, **37**, 35-46.

Maher, M., I. Summersgill (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention*, **28**, 281-296.

McGuigan, D. R. D. (1981). The use of relationships between road accidents and traffic flow in "black spot" identification. *Traffic Engineering and Control*, **22**, 448-453.

Mensah, A. and E. Hauer (1998). Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. *Transportation Research Record*, **1635**, 37-43.

Miaou, S-P., and D. Lord (2003). Modeling traffic crash-flow relationships for intersections. Dispersion parameter, functional form, and Bayes versus empirical Bayes methods. *Transportation Research Record*, **1840**, 31-40.

Miaou, S-P., A. Lu and H. S. Lum (1996). Pitfalls in using R^2 to evaluate goodness of fit of accident prediction models. *Transportation Research Record*, **1542**, 6-13.

Morrison, D.G. and D. C. Schmittlein (1981). Predicting future random events from past performance. *Management Science*, **27**, 1006-1023.

Mountain, L., B. Fawaz and D. Jarrett (1996). Accident prediction models for roads with minor junctions. *Accident Analysis and Prevention*, **28**, 695-707.

Mountain, L., M. Maher and B. Fawaz (1998). The influence of trend on estimates of accidents at junctions. *Accident Analysis and Prevention*, **30**, 641-649.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

OECD road research group (1976). Hazardous road locations: identification and countermeasures. OECD, Paris.

Oh, J.; S. Washington and K. Choi (2004). Development of accident prediction models for rural highway intersections. *Transportation Research Record*, **1897**, 18-27.

Overgaard Madsen, J. C. (2005). *Skadesgradsbasered sortpletudpegning – fra crash prevention til loss reduction i de danske vejbestyrelsers sortpletarbejde*. Ph d afhandling. Trafikforskningsgruppen, Institut for samfundsudvikling og planlægning, Aalborg Universitet, Aalborg.

Partyka, S. (1991). Simple models of fatality trends revisited seven years later. *Accident Analysis and Prevention*, **23**, 423-430.

Persaud, B. and C. Lyon (2007). Empirical Bayes before-after studies: lessons from two decades of experience and future directions. *Accident Analysis and Prevention*, **39** (forthcoming).

Persaud, B., C. Lyon and T. Nguyen (1999). Empirical Bayes procedure for ranking sites for safety investigation by potential for improvement. *Transportation Research Record*, **1665**, 7-12.

Persaud, B., H. McGee, C. Lyon and D. Lord (2003). Development of a procedure for estimating expected safety effects of a contemplated traffic signal installation. *Transportation Research Record*, **1840**, 96-103.

Qin, X., J. N. Ivan, N. Ravishanker and J. Liu (2005). Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov Chain Monte Carlo modelling. *Journal of Transportation Engineering*, **131**, 345-351.

Ragnøy, A., P. Christensen, R. Elvik (2002). *Skadegradstetthet. Et nytt mål på hvor farlig en vegstrekning er*. Rapport 618. Transportøkonomisk institutt, Oslo.

Ragnøy, A. and R. Elvik (2003). *Trafikksikkerhetsanalyse av stamvegnettet*. Rapport 649. Oslo, Transportøkonomisk institutt.

Rosenthal, R. and D. B. Rubin (1978). Interpersonal expectancy effects: the first 345 studies. *Behavioral and Brain Sciences*, **3**, 377-415.

Rothman, K. and S. Greenland (1998). *Modern epidemiology. Second edition*. Lippincott Williams and Wilkins, Philadelphia, PA.

Shankar, V., J. Milton and F. Mannering (1997). Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. *Accident Analysis and Prevention*, **29**, 829-837.

Statens vegvesen. (2006). *Håndbok 115. Analyse av ulykkessteder*. Draft dated October 19, 2006. Vegdirektoratet, Oslo.

Taylor, M. C., A. Baruya and J. V. Kennedy (2002). *The relationship between speed and accidents on rural single-carriageway roads*. TRL report TRL511. Transport Research Laboratory, Crowthorne, Berkshire.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Turner, S. and A. Nicholson (1998). Intersection accident estimation: the role of intersection location and non-collision flows. *Accident Analysis and Prevention*, **30**, 505-517.

Vistisen, D. (2002). *Models and methods for hot spot safety work*. Ph D dissertation. Department for informatics and mathematical models, Technical University of Denmark, Lyngby.

Copyright © Transportøkonomisk institutt, 2007
Denne publikasjonen er vernet i henhold til Åndsverkloven av 1961

Sist utgitte TØI publikasjoner under program: Risikoanalyser og kostnadsberegninger

Realisering av nullvisjonen: Forebygging av fotgjengerulykker og redusering av ulykkenes alvorlighet	889/2007
Nullvisjonen - i teori og praksis	873/2007
Effektkatalog for trafikksikkerhetstiltak	851/2006
Trafikksikkerhetsinspeksjoner: effekter og retningslinjer for god praksis	850/2006
Vegdekkets tilstand og trafikksikkerhet. Betydningen av spordybde, ujevnhet og endringer i tverrfall for ulykkesrisikoen	840/2006
Trafikkstøy i boliger. Virkninger av fasadeisoleringstiltak etter grenseverdiforskriften	836/2006
Økonomisk verdsetting av ikke-markedsgoder i transport.	835/2006
Syklistskader, risiko ved sykling og nyttekostnadsanalyseverktøyet for sykkeltiltak	816/2005
Sykkelulykker. Ulykkestyper, skadekonsekvenser og risikofaktorer.	793/2005
Er bedringen i trafikksikkerheten stoppet opp?	792/2005
Vurdering av behov for halvårlig kontroll av bremser på tunge kjøretøy	790/2005
Etikk og trafikksikkerhetspolitikk	786/2005
Barrierer mot bruk av effektivitetsanalyse i utforming av trafikksikkerhetspolitikk	785/2005
Endring av fartsgrenser. Effekt på kjørefart og ulykker	784/2005
Muligheder og barrierer for trafiksikkerhedsarbejdet i Sverige - en analyse af Vägverket og andre aktører.	759/2005

Visiting and postal address:
Institute of Transport Economics
Gaustadalléen 21
NO 0349 Oslo

Telephone: +47 22 57 38 00
Telefax: +47 22 60 92 00
E-mail: toi@toi.no

www.toi.no

Institute of Transport Economics
Norwegian Centre for Transport Research

- carries out research for the benefit of society and industry
- has a research staff of around 70 highly qualified staff working in various areas of transport research
- co-operates with a number of organisations, research institutes and universities in Norway and in other countries
- carries out high quality research and assessment projects within topics such as traffic safety, public transport, the environment, travel behaviour, tourism, planning, decision-making processes, transport economics, and freight
- publishes research findings in the Centre's own report series, on the Internet, in the periodical "Samferdsel", as well as in national and international scientific journals and publications
- participates in the Oslo Centre for Interdisciplinary Environmental and Social Research (CIENS) located near the University of Oslo