Safety Science 48 (2010) 1111–1122

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

# On prediction in road safety

Ezra Hauer *

*35 Merton Street, Toronto, Ontario, Canada M4S 3G4*

ARTICLE INFO

*Article history:*
Received 12 September 2009
Received in revised form 2 March 2010
Accepted 3 March 2010

*Keywords:*
Prediction
Road safety
Safety targets
Evaluation

ABSTRACT

Prediction is about potential outcomes: *what will happen if* and *what would have happened if*. The first question arises when safety targets are set, the second when the effect of an intervention on safety is to be evaluated. There are many ways to predict. For the same data different prediction methods produce different predictions. What targets are set and what estimates of intervention effect are produced will depend on what method of prediction is chosen. Therefore one has to determine what method tends to predict best. To do so empirically one asks what method would have predicted best had it been applied in the past and then one assumes, inductively, that the same would apply in the future. Quantitative measures of prediction quality are suggested and it is shown how these measures of prediction quality allow one to determine which of two prediction methods should be preferred.

The suggested approach was applied to two data sets: The time series of motor vehicle accident fatalities in Province A and in Province B. On the basis of this analysis one may draw tentative conclusions for these jurisdictions and the methods tested; one can say what method seems preferable, what is the average size of bias than needs to be corrected and how accurate is the prediction likely to be. Broader conclusions will emerge once many additional methods of prediction are applied to data from many other jurisdictions and pertaining to a variety of circumstances.



## 1. Introduction

Safety predictions are needed for two main purposes:

1. To set policy or program targets. For this purpose one needs to predict what *would be* the future safety of the unit[1] *if* the policy or programs will not implemented. Safety targets are then set on this basis.
2. To estimate what was the effect a policy, program, or treatment on the safety of the unit. For this purpose too one needs to predict what *would have been* the safety of the unit *had* the policy, program or treatment not been implemented.

The difference between the two circumstances is one of timing: for target setting the prediction is produced before the program or policy is implemented whereas for effect estimation the prediction is produced after the intervention was implemented.[2]

There are many commonly used methods of prediction. Some methods rely on the extrapolation of past trends; some make use comparison groups, another family of methods attempts to model the causal factors that govern the evolution of the safety of a unit. In this paper the question is not how to best describe the structure underlying a time series of data. The question is which method predicts best. This is why the emphasis is not on theory-motivated time series analysis methods and the characterization of their statistical properties. The emphasis is inductive and empirical; a method is attractive for future use if, had it been used in the past it would have predicted better than other methods. From this perspective the theoretical underpinnings of a method and the statistical properties of its fit do not matter.

The quality of a method for safety prediction depends on the nature of the unit for which the prediction is prepared. That is, that one method of prediction may be best when the unit is an intersection or road section while a different method may be best when the unit is a province, state, or country. It is also possible that one method is best for making predictions about the next few years while another when predictions into the more distant future is of interest. In this paper the focus is on predictions when the unit is a province, state or country. In this context predictions for both the near and the more distant future are of interest. Near future predictions are needed to evaluate the effect of interventions in the short term; longer term predictions are needed to tell what might be the safety of some jurisdiction if past practices

* Tel.: +1 416 483 4452.
*E-mail address:* Ezra.Hauer@Utoronto.ca

[1] Policies, programs, and treatments apply to units. A unit can be a state, a region, a city, a road section, drivers of an age cohort, trucks of some type, etc.

[2] This has implications about the kind of information available at the time of the prediction is made. For target setting predictions one can only forecast what are likely to be the traits of the unit during the period to which the prediction applies. In contrast, since for effect evaluation the period of prediction is in the past, some such traits can be known.

0925-7535/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2010.03.003

were to continue and no novel programmes or new policies implemented.

In this paper the approach to answering the question of which method predicts best is empirical. Available historical data will be used to ask how a method of prediction would have performed had it been used in the past. For clarity it is best to define the differences between Prediction, Forecast, and Estimate. *Predictions* are about a state of the world that did not and will not exist; they pertain to 'potential outcomes'.[3] Thus, prediction is always associated with a conditional clause such as 'what would be if' or 'what would have been if'. Should it happen that, in the event, the policy, program, or intervention will not be implemented then, what was a prediction, turns into a *forecast*. Unlike a prediction, a forecast is about a state of the world that will exist and therefore the quality of a forecast can be judged by juxtaposing it to what eventually materialized. In contrast to forecasts which pertain to events that will materialize in the future, *estimates* are about events in the past. Thus, e.g., we can use the history of accident occurrence to estimate the expected number of accidents.

The quality of estimates is the subject of much statistical theory and may be regarded well known. Similarly, there exists a very large body of knowledge about forecasting and forecast quality. This lore is embodied in many books and learned journals.[4] The quality of 'potential outcome predictions'[5] is less well explored.

## 2. Previous work

The issue of prediction in road safety was examined explicitly in Hauer et al. (1991). Using a time series of annual fatal and injury accidents in each of 10 Canadian provinces, predictions were formed by four methods: (a) using 1 year's count to predict for the next, (b) using the average for 2 years to predict the for the next, (c) using the least-squares fit to three counts to predict the next and, (d) using comparison group, i.e. multiplying the count in Province A in year $y$ by the ratio of counts Province B for years $y+1$ and $y$ to predict the expected accident count in Province A in year $y+1$. It turned out that, in spite of using more data than method 'a', methods 'b' and 'c' gave worse predictions. Method 'd' gave good predictions if the comparison group had many accidents. We found that "one can predict better not using a comparison group than using one that is too small" (p. 599). In spite of common belief in the importance of similarity between the 'treatment' and the 'comparison group', it turned out that similarity played only a minor role. The issue of predicting by method 'd' was analysed in Hauer (1991).

A more thorough examination of prediction in road safety is in the doctoral thesis of Quaye (1992). Much of what follows draws on his work. Quaye's object was to propose means to measure the performance of prediction methods and, on this basis, to ascertain whether one method is better than another. Quaye made a distinction between prediction methods that extrapolate a time series of accident counts and between econometric methods that make use of information about some causal variables. For extrapolation methods his data were the fatality counts of Canadian provinces; for the econometric models he used also data about population, vehicle registration, and vehicle-kilometres of travel ($V_{kmt}$). Some of his results were published in Quaye and Hauer (1993) and presented in Quaye and Hauer (1994). In these we found, e.g., that using twelve different methods to predict the number of fatal accidents in Province A it seemed that extrapolation of a quadratic ordinary least-squares fit produced better predictions that the averaging of past '$y$' accident counts and was also better than various econometric models that used data about population, registered vehicles and vehicle-kilometres of travel. In 1994 the trail seems to have grown cold. No further research into the quality of road safety predictions as produced by various prediction methods was found.

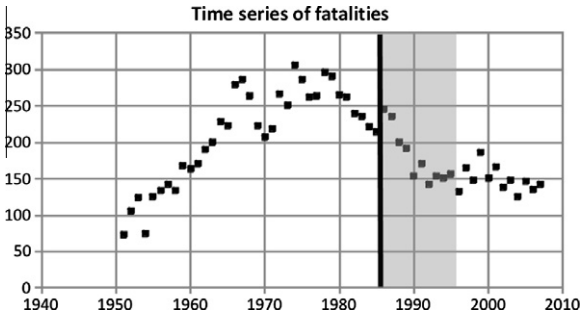


**Fig. 1.** Count of road accident fatalities in Province A.

## 3. Same data, different predictions

The purpose of this section is to show that methods of prediction are many and that these produce a multitude of diverse predictions. In the face of this multitude and diversity one asks: "Which method of prediction should one believe more?" To explain the issues arising from this question in a tangible setting data about the number of persons killed in crashes in Province A (Fig. 1) will be used.

This mountain-like shape depicts the evolution of a process over time.[6] What forces shaped this process can be said only in general terms: they are the changes in population, the processes of urbanization and motorization, advances in medicine, improved roads, safer vehicles, changes attitude to risk and in other societal norms, etc.

Suppose that we are at the end of 1985 and wish to predict what would be the number of fatalities for the next 10 years if the same forces continued to exert their influence and no substantially new programs and policies were implemented.[7] The predictions for 1986–1995 are to be based on data for the years to the left of the solid line; what is to the right of the solid line is at this point in time unknown.

Prediction methods are many. As the purpose here is illustration, only relatively simple methods of prediction will be used. Thus, e.g., one could extrapolate the two most recent observed counts, or fit an ordinary least squares regression (OLS) a line to the $n$ (=3, 4, 5, ...) most recent observed counts, or fit a polynomial

[3] The 'potential outcome' terminology favoured by Rubin (see, e.g., Rubin, 2005) is used instead of the equivalent expression: 'counterfactual'.

[4] One list (http://www.forecastingprinciples.com/books1.html) contains entries for 75 books on forecasting, The University of Toronto libraries presently subscribe to eight journal that have 'forecasting' in their title. Most such books and journals are oriented towards business and economics.

[5] In the context of safety evaluation the word 'prediction' is a poor choice. Evaluations are undertaken after an intervention was implemented, and therefore 'prediction', as used in this context, refers to a time and state of affairs which is already in the past at the time when the evaluation commences. This is awkward because in common usage prediction refers to events in the future. It might have been better to use 'postdiction' or 'retrodiction' instead. However, both terms are already in use and have acquired specialized meanings. Besides, the word prediction has been used in the sense defined here earlier (Hauer, 1997) and coining a new term might lead to confusion. In the rest of this paper the word 'prediction' will be used instead of the longhand expression: 'potential outcome safety prediction'.

[6] The process evolved in a similar way in nearly all developed countries.

[7] This is similar to what Broughton et al. (2000) aimed for when preparing the framework for setting the 2010 Target in the UK. Their goal was to predict "what would be expected if there were no further DESS measures and only the core road safety activities were undertaken (at the 1998 level of effect) during the period to 2010; this was done by extrapolating trends from the 1983–1998 period." (Broughton, 2006). The acronym DESS stands for Drink/driving, Engineering, Secondary Safety improvements to vehicles.

**Table 1**
A variety of predictions.

| Method and variant | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS linear, $n = 2$ | 207.0 | 200.0 | 193.0 | 186.0 | 179.0 | 172.0 | 165.0 | 158.0 | 151.0 | 144.0 |
| OLS linear, $n = 3$ | 202.3 | 191.8 | 181.3 | 170.8 | 160.3 | 149.8 | 139.3 | 128.8 | 118.3 | 107.8 |
| OLS linear, $n = 5$ | 199.9 | 188.4 | 176.9 | 165.4 | 153.9 | 142.4 | 130.9 | 119.4 | 107.9 | 96.4 |
| OLS linear, $n = 10$ | 215.0 | 207.7 | 200.5 | 193.2 | 185.9 | 178.6 | 171.4 | 164.1 | 156.8 | 149.5 |
| OLS Quadratic, $n = 3$ | 214.0 | 221.0 | 235.0 | 256.0 | 284.0 | 319.0 | 361.0 | 410.0 | 466.0 | 529.0 |
| OLS Quadratic, $n = 5$ | 210.4 | 209.4 | 211.4 | 216.4 | 224.4 | 235.4 | 249.4 | 266.4 | 286.4 | 309.4 |
| OLS Quadratic, $n = 10$ | 181.9 | 156.6 | 128.3 | 96.9 | 62.6 | 25.2 | -15.1 | -58.5 | -104.8 | -154.2 |
| Holt smoothing,[a] $A$[b] | 207.0 | 199.6 | 192.2 | 184.8 | 177.5 | 170.1 | 162.7 | 155.4 | 148.0 | 140.6 |
| Holt smoothing, $B$[c] | 204.4 | 200.7 | 190.5 | 203.7 | 194.8 | 188.5 | 175.2 | 163.3 | 151.2 | 131.3 |
| Hoerl function, $n = 5$ | 201.2 | 188.4 | 175.3 | 162.2 | 149.3 | 136.8 | 124.8 | 113.4 | 102.6 | 92.6 |
| Hoerl function, $n = 10$ | 210.9 | 202.3 | 193.4 | 184.5 | 175.6 | 166.7 | 158.0 | 149.4 | 141.0 | 132.9 |
| Hoerl function, $n = 20$ | 218.4 | 212.5 | 206.6 | 200.7 | 194.7 | 188.7 | 182.7 | 176.7 | 170.8 | 164.9 |
| Oppe–Koornstra | 256.7 | 254.7 | 252.5 | 250.1 | 247.4 | 244.6 | 241.6 | 238.4 | 235.2 | 231.8 |
| What materialized | 245 | 236 | 200 | 192 | 154 | 170 | 143 | 153 | 151 | 157 |

[a] Let $y_i$ be the accident count at time $i$ ($i = 1, 2, \ldots, n$) and $s_i$ the 'smoothed' value. Holt smoothing consists of the recursive relationships $s_i = \alpha y_i + (1 - \alpha)(s_{i-1} + u_{i-1})$ and $u_i = \gamma(s_i - s_{i-1}) + (1 - \gamma)u_{i-1}$ in which there are two smoothing constants ($\alpha$ and $\gamma$) and starting values ($s_0$ and $u_0$).

[b] In variant $A$ the smoothing and starting values were selected so as to minimize the sum of one-step-ahead absolute forecast errors. The prediction in Table 1 for years $Y > 1985$ is $s_{1985} + (Y - 1985) \times u_{1985}$. For the Province A data, 214.3 + (Year − 1985) × (−7.37).

[c] In variant $B$, the $\alpha$, $\gamma$, $s_0$ and $u_0$ for $n$-steps-ahead predictions are estimated so as to minimize the sum of $n$-step-ahead absolute prediction errors.

to these points, or use double exponential smoothing, or fit a Hoerl function or, following Oppe and Koornstra (1990) and Oppe (1991), extrapolate risk and exposure and predict using their product.[8,9] The resulting predictions are in Table 1.

The juxtaposition of these predictions and the accident counts that materialized during 1986–1995 (the black squares) is in Figs. 2–5.

One should resist the temptation to form conclusions on the basis of an illustration based on few simple methods of prediction that rely on data from only one jurisdiction (Province A), and use predictions for only one point in time (1985). The few rows of Table 1 are just a smattering of what could be listed. Instead of OLS one could use maximum likelihood, instead of linear or quadratic polynomial regressions one could choose from many function families, in addition to the few $n$'s listed many others could be used, instead of Holt smoothing one could use one of the many autoregressive or state space models for time series data etc. It is obvious that as prediction methods are many and each method has many variants, predictions can be indeed very many. Furthermore, as in Table 1, the many predictions usually vary widely. Naturally one asks which method and variant are best in what circumstance. 'Circumstance' may refer to type of jurisdiction, the shape of the time series profile, whether the prediction is for but few years or for further into the future and so on. Determining which approach to prediction is best requires substantive empirical inquiry involving a juxtaposition of what would have been predicted and what has materialized for several jurisdictions and points in time. Before embarking on this task for two jurisdictions, a brief a discussion of two topics is in order: the issue of extrapolation versus causal modeling in prediction, and the question of how the quality of predictions can to be measured.

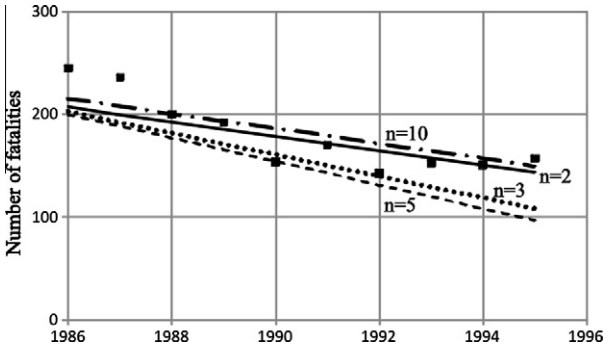


**Fig. 2.** Prediction by straight lines fitted to the last $n$ data points.

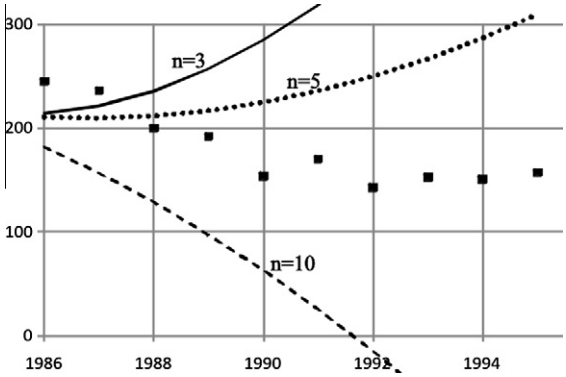


**Fig. 3.** Predictions by extrapolating a quadratic polynomial fitted to the last $n$ data points.

## 4. Approaches to prediction: extrapolation versus causation

All the predictions in Table 1 rely on extrapolation. Even the Oppe–Koornstra method relies on the extrapolation of a time series of risks and another extrapolation of a time series of exposures. Extrapolation rests on the assumption that time-trends of the past will continue into the future. It is commonly believed that one would predict better if the causes of the evolution of a process

[8] The Oppe–Koornstra approach entails two equations. The first traces the evolution of risk over time ($R_t = \exp(\alpha t + \beta)$ in which $R_t$ is the risk at time $t$, with $\alpha$ and $\beta$ as parameters to be estimated from data). The second traces the evolution of exposure over time ($V_t = V_m/\{1 + \exp[-(at + b)]\}$ in which $V_t$ represents the vehicle-kilometres of travel at time $t$, $V_m$ is the maximum (saturation) $V_{kmt}$ with $a$ and $b$ parameters estimate from data). The number of accidents is the product $R_t \times V_t$.

[9] One could also use econometric models making use of variables such a population, income, alcohol, and consumption. Ranging from simple multivariable regressions to complex structural models. However, models of this kind will have difficulty replicating the fatality peak evident in Fig. 1 and will require the making predictions for all variables in the model.

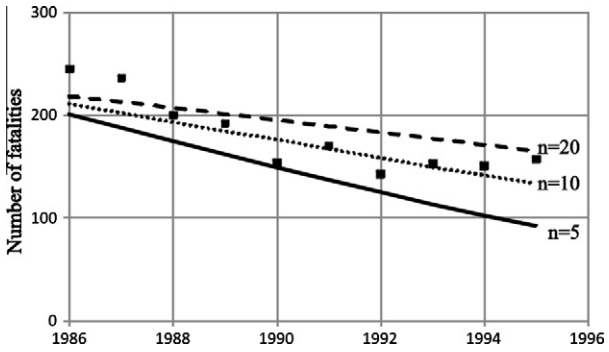


**Fig. 4.** Predictions by extrapolating a Hoerl function fitted to the last $n$ data points.

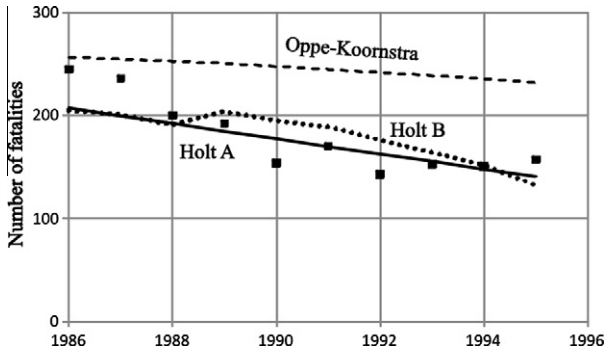


**Fig. 5.** Two variants of Holt smoothing and the Oppe–Koornstra prediction.

were understood and the magnitudes of these causes were known or could be predicted. Fig. 6 is an attempt to clarify.

The cause of the change in the sequence of accident counts is not the passage of time; accidents depend causally on various factors. Thus, e.g., the number of traffic fatalities in a province depends on the number of drivers, amount of pedestrian traffic, quality of health care, norms of behaviour, on what roads and in what environment travel takes place, etc. These factors themselves change with time as indicated by the top grey arrow.

Prediction by extrapolation assumes that causal factors will change with time is such a way that the future is a smooth continuation of past trends. This is represented by the curved dashed arrow on the right which bypasses the box containing causal factors. However, if one could know how expected accidents change under the influence of change in causal factors and if the value of these causal factors in the period for which the prediction is needed is could be known, estimated or predicted, then one could do away with the assumption that the future is an unspecified continuation of the past. The hope is that replacing the belief in an unspecified 'continuity of trends' by information about causal factors and knowledge about cause–effect relationship might improve prediction. This is represented by the looping solid arrow on the left.

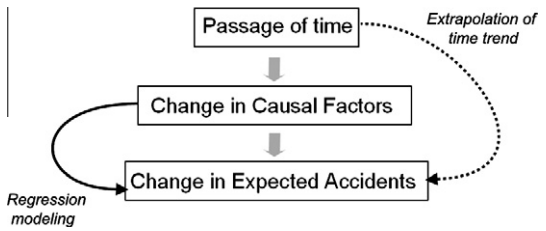


**Fig. 6.** Causal versus extrapolation of time-trend prediction.

If data about the past values of all important causal factors were available, if the functional relationship between expected accident and causal factors was sufficiently understood, and if regression modeling was up to the task of distinguishing between association and causation, then it might be possible to describe how the expected number of accidents would change if the values of some causal factors changed. At present, none of these conditions fully holds. Furthermore, to predict what would be the expected number of accidents at some future time one would also have to be able to predict the magnitudes of all the causal factors for that future time, further complicating the task of prediction by causal models. For these reasons it is not clear whether predictions based on the use of data in statistical regression (structural) models will usually be better than those based on extrapolating time-trends.[10]

In the present context of prediction for large jurisdictions there is an additional reason for shunning causal models. As in Fig. 1, in most large jurisdictions there is a period during which fatalities increased, reached a peak and followed by a period of decline and stabilization. All the while causal variables such a population, exposure etc. increased. For a satisfactory explanation of the peak in the time series of fatalities one has to include in the model a cause–effect representation of how risk (≡accidents per unit of exposure) changes as a function of causal variables (such as changes in public mores, investment in infrastructure, medicine, and urbanization) At present, knowledge of this kind is insufficient. Even if it was known how risk depends on, say, public mores or the extent of urbanization, it still would be necessary to predict what these will be in the future. Unless the process of the evolution of public mores or of urbanization could be described as a function of some variables other than time, extrapolation would still be needed. The hope that the evolution of fatalities can be explained by cause and effect, without resorting to extrapolation of time-trends leads to infinite regress. For these reasons, in the present inquiry, only time-based extrapolation approaches will be examined.

## 5. Prediction quality

In Section 3 the story was that we are at the end of 1985 and need to predict what would be the number of fatalities in Province A for 1986–1995 if no substantially new policy or program were to be implemented in those 10 years. If we predict by, say, linear extrapolation of an OLS fit with $n = 5$, then the predictions are those in row 4 of Table 1 and repeated below in row $P$ of Table 2. The count of fatalities that materialized in that period is in row $X$. The differences $X - P$ between the two are in row $D$. In row $d$ is the relative difference, the difference as proportion of the count that materialized. Thus e.g., if $d \equiv (X - P)/X = 0.18$, then the predicted value is 18% below the accident count.

Were one to predict similarly at the end of 1984, one would obtain Table 3.

As the first year of data is for 1952 and when $n = 5$, the first such table could be produced at the end of 1956. Since the last year for which we have fatality counts is 2007, and each table is for 10 'steps ahead' years, the last table that can be produced is for 1997. Thus we can have 42 such tables and 42 rows of relative differences for $1, 2, \ldots, 10$ steps-ahead predictions. Selected rows from this composite table are in Table 4.

[10] In the community of forecasters there is vigorous debate about the quality of forecasts obtainable by the relatively simple methods of time-series extrapolation and forecasts produced by more complex regression models that make use of important causal variables and often are often based on some theory which takes the form of simultaneous equations. The weight of the evidence at this time suggests that complex regression models do not necessarily produce better forecasts than simpler univariate time-trend extrapolations. As Makridakis and Hibon (2000) say: "Statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones." p. 452.

The interpretation of this table is as follows. Suppose that using the count of fatalities in Fig. 1 one was to predict what will be the annual number of fatalities in the next 10 years by extrapolating the ordinary-least-squares fit to the last 5 years of fatality counts. A one-step-ahead prediction (for 1957) when the last year of data is 1956 would prove to have $d = -0.11$ while a seven step ahead prediction (for 1991) when the last year of data is 1984 would prove to have $d = 0.17$. It should be clear that the entries in the table are not statistically independent. Thus, e.g., for $s = 1$ the $d$ in 1956 ($-0.11$) and the $d$ in 1957 (0.03) share the same four of five data points to estimate their prediction equations. The last three rows of the table contain summary statistics. Judging by $\bar{d}$, the predictions (the $P$'s) for the years 1956–1997 will be on the average about 1% higher[11] than the count of fatalities (the $X$'s)when $s = 1$, 2 and 3 while for $s = 9$ or 10 the $P$'s would be, on the average, 6% higher than $X$'s. Here the bias ($\bar{d}$) increases with the number of steps ahead. The standard error of the $d$'s around $\bar{d}$ is in the before last row of Table 4. The further into the future one predicts the larger is the standard error of predictions. Thus, e.g., when predicting (by OLS with $n = 5$) 6 years into the future one should expect the average prediction to be 2% too high and its standard error to be ±39%. As is evident, the standard error of prediction increases with $s$.

In sum, the quality of a method of predictions can be judged by two attributes. First, one would like predictions to be unbiased in the sense that when the prediction is $P$ then the expected value of the corresponding $X$'s is also $P$. Second, one would like predictions to be 'close' to the expected value of $X$. The notion of being 'close' can be variously defined. One common measure of closeness is the standard error (SE); another is the mean absolute error (MAE). While many measures of prediction error are in use[12], in what follows the measures of prediction quality will be the *bias*, as measured by the average of $d$'s (the $\bar{d}$ or $d$-bar), and the *standard error* of the $d$'s denoted by $SE_{d-bar}$. The attributes of being unbiased and that of closeness are not independent. Methods that produce predictions with significant bias will tend to also have large SE and MAE. Therefore in many cases one can combine them into one measure of performance, the standard error around 0. This measure, to be denoted by $SE_0$ is defined by $SE_0 = (SE_{d-bar}^2 + bias^2)^{0.5}$. Having created the yardstick for measuring prediction quality we can now turn to the task of comparing methods and their variants in terms of the quality of predictions that they produce.

## 6. Comparing variants and methods

The central question is which of two prediction methods and variants would have predicted better. Thus, e.g., had we predicted by the linear extrapolation of an OLS fit, would we have predicted better with $n = 5$ or with $n = 10$? Having specified what will measure the quality of predictions an answer can now be attempted. The raw material is in Table 5 the top half of which was taken from the summary rows of Table 4 and the bottom half of which was similarly generated except that the extrapolation was of an OLS fit to the last ten fatality counts ($n = 10$).

The three comparisons are in the Figs. 7–9.

The negative bias is larger with $n = 10$ than with $n = 5$ for all $s$. On the other hand, with the exception of $s = 1$, when $n = 10$ the predictions have a smaller standard error (around $\bar{d}$ and around 0) than when $n = 5$. One may conclude that when $s = 1$ it would have been better to use $n = 5$ while $n = 10$ would have been preferred for $s > 1$. The predictions would have been better, on the average,[13] had they been reduced by $\bar{d}$. Thus, e.g., if $s = 5$ then the predictions obtained by linear extrapolation should have been reduced by 5% (as per Table 5.)

The method that allowed the comparison of $n = 5$ and $n = 10$ can be used to compare any two $n$'s. How the three measures of prediction quality vary with $s$ and $n$ is shown in the three figures below.

The bias (Fig. 10) increases with $s$ and $n$. The standard error around $\bar{d}$ is smallest between $n = 9$ and 10. The standard error around 0 is also smallest between $n = 9$ and 10. It follows that if

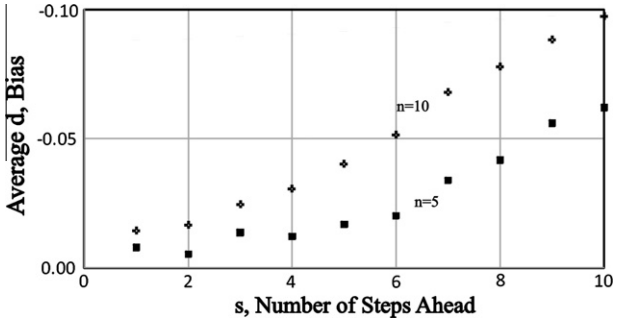


**Fig. 7.** Bias as a function of $s$ for $n = 5$ and $n = 10$.

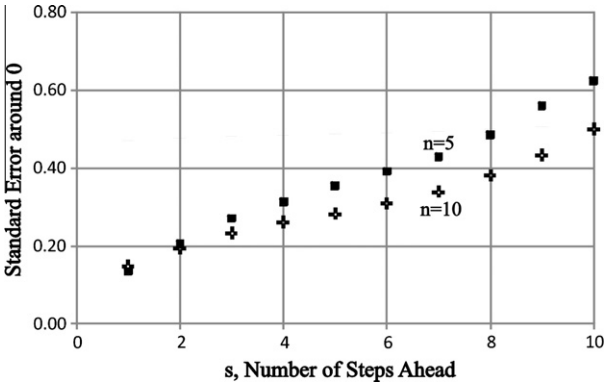


**Fig. 8.** Standard error around $\bar{d}$ as a function of $s$ for $n = 5$ and $n = 10$.

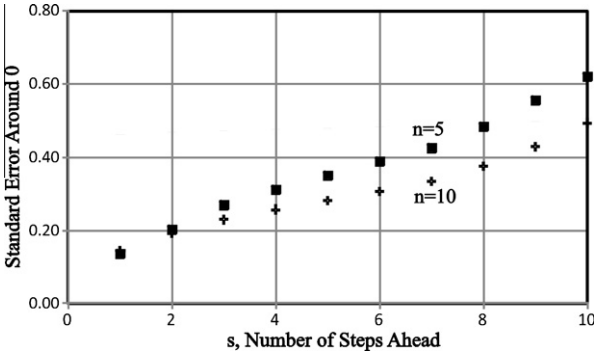


**Fig. 9.** Standard error around 0 as a function of $s$ for $n = 5$ and $n = 10$.

[11] When $P > X$ then $D$ (and also $d$) are negative. A negative $d$ (or $\bar{d}$) indicates the predictions ($P$) are to high. When a linear prediction equation is used $P$ will tend to be too high when the underlying process is a downward bending curve (the second derivative is negative).

[12] Common measures of forecast accuracy are also the MAPE (Mean Absolute Percentage Error), MdAPE (Median Absolute Percentage Error), sMAPE (Symmetric Mean Absolute Percentage Error), sMdAPE (Symmetric Median Absolute Percentage Error), MdRAE (Median Relative Absolute Error), GMRAE (Geometric Mean Relative Absolute Error), and the MASE (Mean Absolute Scaled Error). From Hyndman and Koehler (2006).

[13] $\bar{d}$ is the right correction if it is not clear whether, at the time when the prediction is made, that the underlying process is non-linear. If at the time the prediction is made it be clear that the underlying process is upward or downward bending then the use of a linear predictor does not make much sense.

**Table 2**
Predictions for the 10 years following 1985.

| | Years | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $s$ | Steps ahead | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $P$ | OLS linear, $n = 5$ | 199.9 | 188.4 | 176.9 | 165.4 | 153.9 | 142.4 | 130.9 | 119.4 | 107.9 | 96.4 |
| $X$ | Count of fatalities | 245 | 236 | 200 | 192 | 154 | 170 | 143 | 153 | 151 | 157 |
| $D$ | Differences | 45.1 | 47.6 | 23.1 | 26.6 | 0.1 | 27.6 | 12.1 | 33.6 | 43.1 | 60.6 |
| $d$ | Rel. diff. $D/X$ | 0.18 | 0.20 | 0.12 | 0.14 | 0.00 | 0.16 | 0.08 | 0.22 | 0.29 | 0.39 |

**Table 3**
Predictions for the 10 years following 1984.

| | Years | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $s$ | Steps ahead | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $P$ | OLS linear, $n = 5$ | 210.1 | 198.6 | 187.1 | 175.6 | 164.1 | 152.6 | 141.1 | 129.6 | 118.1 | 106.6 |
| $X$ | Count of fatalities | 214 | 245 | 236 | 200 | 192 | 154 | 170 | 143 | 153 | 151 |
| $D$ | Differences | 3.9 | 46.4 | 48.9 | 24.4 | 27.9 | 1.4 | 28.9 | 13.4 | 34.9 | 44.4 |
| 2 | Rel. diff. $D/X$ | 0.02 | 0.19 | 0.21 | 0.12 | 0.15 | 0.01 | 0.17 | 0.09 | 0.23 | 0.29 |

**Table 4**
Relative differences [$d \equiv (X - P)/X$] by 'year' and 'steps ahead' with OLS, linear, $n = 5$.

| Last year of data | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1956 | −0.11 | 0.05 | −0.03 | −0.05 | 0.01 | 0.01 | 0.09 | 0.02 | 0.18 | 0.17 |
| 1957 | 0.03 | −0.08 | −0.12 | −0.07 | −0.09 | −0.01 | −0.10 | 0.07 | 0.05 | −0.09 |
| 1958 | −0.02 | −0.02 | 0.04 | 0.04 | 0.12 | 0.06 | 0.22 | 0.21 | 0.11 | −0.09 |
| ⋮ | | | | | | | | | | |
| 1980 | −0.09 | −0.20 | −0.24 | −0.33 | −0.39 | −0.22 | −0.28 | −0.53 | −0.61 | −1.02 |
| 1981 | −0.10 | −0.11 | −0.17 | −0.19 | −0.02 | −0.05 | −0.22 | −0.25 | −0.54 | −0.37 |
| 1982 | 0.03 | 0.03 | 0.06 | 0.24 | 0.27 | 0.21 | 0.25 | 0.15 | 0.31 | 0.28 |
| 1983 | 0.01 | 0.04 | 0.22 | 0.25 | 0.18 | 0.22 | 0.11 | 0.27 | 0.23 | 0.37 |
| 1984 | 0.02 | 0.19 | 0.21 | 0.12 | 0.15 | 0.01 | 0.17 | 0.09 | 0.23 | 0.29 |
| 1985 | 0.18 | 0.20 | 0.12 | 0.14 | 0.00 | 0.16 | 0.08 | 0.22 | 0.29 | 0.39 |
| ⋮ | | | | | | | | | | |
| 1995 | −0.12 | 0.11 | 0.01 | 0.23 | 0.06 | 0.15 | 0.00 | 0.08 | −0.07 | 0.09 |
| 1996 | 0.14 | 0.05 | 0.25 | 0.09 | 0.18 | 0.02 | 0.10 | −0.04 | 0.12 | 0.06 |
| 1997 | −0.04 | 0.17 | −0.03 | 0.06 | −0.13 | −0.06 | −0.25 | −0.07 | −0.16 | −0.11 |
| $\bar{d}$ = average $d$ (bias) | −0.01 | −0.01 | −0.01 | −0.02 | −0.02 | −0.02 | −0.03 | −0.04 | −0.06 | −0.06 |
| Standard error around $\bar{d}$ | 0.13 | 0.20 | 0.27 | 0.31 | 0.35 | 0.39 | 0.42 | 0.48 | 0.55 | 0.61 |
| Standard error around 0 | 0.13 | 0.20 | 0.27 | 0.31 | 0.35 | 0.39 | 0.42 | 0.48 | 0.55 | 0.62 |

**Table 5**
Statistics for $n = 5$ and $n = 10$.

| $n$ | | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | $\bar{d}$ = average $d$ (bias) | −0.01 | −0.01 | −0.01 | −0.02 | −0.02 | −0.02 | −0.03 | −0.04 | −0.06 | −0.06 |
| | Standard error around $\bar{d}$ | 0.13 | 0.20 | 0.27 | 0.31 | 0.35 | 0.39 | 0.42 | 0.48 | 0.55 | 0.61 |
| | Standard error around 0 | 0.13 | 0.20 | 0.27 | 0.31 | 0.35 | 0.39 | 0.42 | 0.48 | 0.55 | 0.62 |
| 10 | $\bar{d}$ = average $d$ (bias) | −0.02 | −0.02 | −0.03 | −0.03 | −0.05 | −0.06 | −0.07 | −0.08 | −0.09 | −0.10 |
| | Standard error around $\bar{d}$ | 0.15 | 0.19 | 0.23 | 0.26 | 0.28 | 0.30 | 0.33 | 0.37 | 0.42 | 0.49 |
| | Standard error around 0 | 0.15 | 0.19 | 0.23 | 0.26 | 0.28 | 0.31 | 0.34 | 0.38 | 0.43 | 0.50 |

for Province A one was to predict using the extrapolation of a linear OLS fit, it would be best to do so using an $n$ of about10. (In general there is no reason to expect that the same $n$ is best for all $s$. In the present case, however, the trough[14] of the function is not a function of $s$.) The performance of such predictions for $n = 10$ and $s = 1$–10 was shown in the lower part of Table 5.

The merit of the suggested approach is now evident. First, it can answer questions of practical interest.[15] Second, it provides an estimate of the average bias inherent in the prediction method and thereby a way for removing it. Third it produces straightforward

[14] While the process underlying Fig. 1 is not a straight line it can be locally approximated by one. For the approximation to be 'local', $n$ has to be small. However, for the slope of the regression line to be reliably estimated n must not be too small. The trough in Figs. 11 and 12 is the result of these two opposite considerations.

[15] Here the question was how many data points should be used in fitting the straight line used for producing predictions.

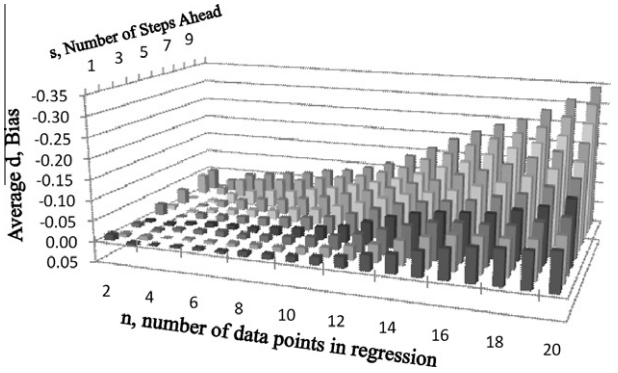


**Fig. 10.** Linear prediction: bias as a function of *s* and *n*.

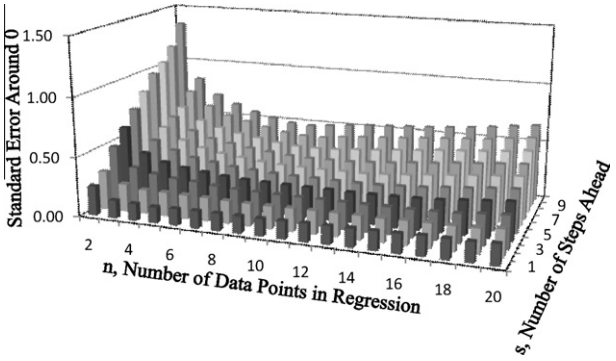


**Fig. 12.** Linear prediction: standard error around 0 as a function of *s* and *n*.

and realistic estimates of the standard error of the predictions for every 's' (such as those in the two last rows of Table 5).

Having established what quality of prediction is attainable in Province A by the extrapolation of a linear OLS fit, the next question is whether a different method of prediction would do better. As noted in Section 2, Quaye found that the extrapolation of a quadratic OLS fit produced good predictions. Therefore we will attempt to determine whether in Province A the quadratic OLS fit usually out-performs the linear OLS fit (in spite of the impression from Fig. 3).

Fig. 13 for the quadratic prediction corresponds to Fig. 10 for the linear one. The main difference is that here the bias is mostly positive (*P* is too low) whereas with the linear prediction the bias is mostly negative (*P* was too high). This difference is due to the shape of the 'mountain' in Fig. 1. With both methods the bias tends to increase with *s* and *n*.

Fig. 14 for the quadratic prediction corresponds to Fig. 11 for the linear one. For ease of comparison the vertical scale has been truncated so as to show the same portion of the function in both figures. It is now clear that for Province A the standard error of the quadratic predictions is always larger than that for the linear one.

Fig. 15 corresponds to Fig. 12 and their comparison leads to the same conclusion: for fatalities in Province A predicting by the extrapolation of a straight line OLS fit would have predicted better than using a quadratic polynomial. Once again the suggested approach for measuring the quality of predictions helps to answer the question of interest straightforwardly.

In this section we established the procedure for determining which variant of a method to use under what circumstances and which of two methods would have usually given better predictions. It should be noted that what has been concluded so far pertains to a specific data set, the time series of fatalities in Province A.

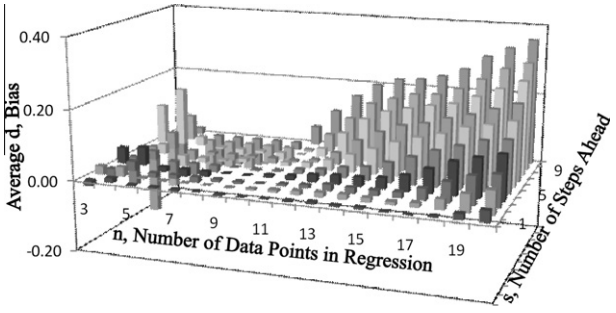


**Fig. 13.** Quadratic prediction: bias as a function of *s* and *n*.

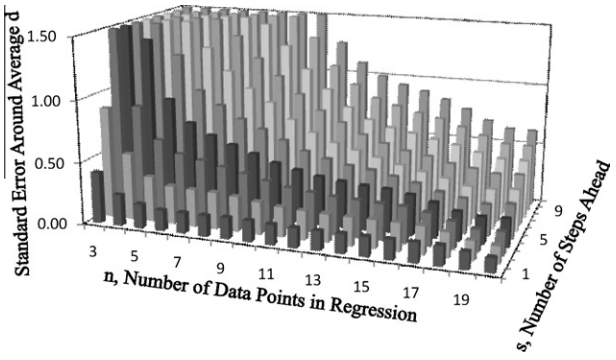


**Fig. 14.** Quadratic prediction: standard error around $\bar{d}$ as a function of *s* and *n*.

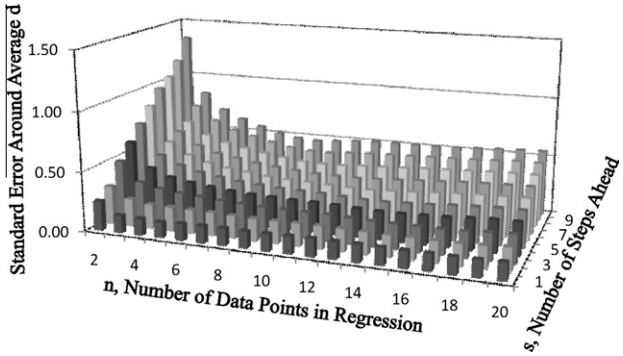


**Fig. 11.** Linear prediction: standard error around $\bar{d}$ as a function of *s* and *n*.

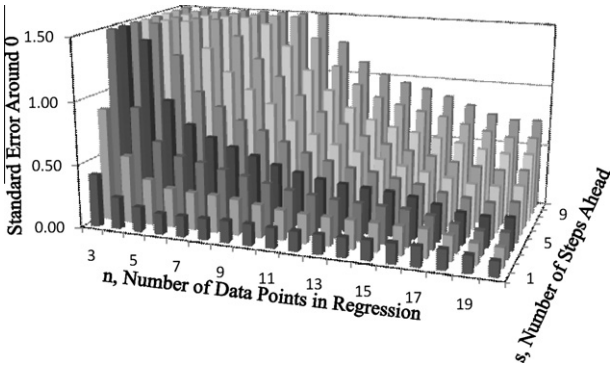


**Fig. 15.** Quadratic prediction: standard error around 0 as a function of *s* and *n*.

**Table 6**
Measures of performance for Holt smoothing A.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | −0.01 | −0.02 | −0.02 | −0.03 | −0.04 | −0.05 | −0.06 | −0.07 | −0.09 | −0.10 |
| Standard error around $\bar{d}$ | 0.12 | 0.17 | 0.23 | 0.26 | 0.28 | 0.30 | 0.33 | 0.36 | 0.42 | 0.46 |
| Standard error around 0 | 0.13 | 0.17 | 0.23 | 0.26 | 0.29 | 0.31 | 0.33 | 0.36 | 0.43 | 0.47 |

**Table 7**
Measures of performance for Holt smoothing B.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | −0.01 | −0.04 | −0.02 | −0.05 | −0.03 | −0.05 | −0.05 | −0.10 | −0.07 | −0.11 |
| Standard error around $\bar{d}$ | 0.12 | 0.17 | 0.21 | 0.28 | 0.29 | 0.34 | 0.38 | 0.47 | 0.52 | 0.66 |
| Standard error around 0 | 0.13 | 0.17 | 0.21 | 0.28 | 0.30 | 0.34 | 0.38 | 0.48 | 0.53 | 0.67 |

Whether similar conclusions hold for other jurisdictions can be established only by performing similar analyses for data from many other jurisdictions.

The following two sections are best regarded as more complete illustration of the approach to determining which method tends to predict best. The illustration is still limited to the simple methods first introduced in Table 1 and there is no intent to claim that other prediction methods would not perform better. In the next section we return to the Province A data to determine whether the other prediction methods introduced earlier in Section 3 can do better than the linear prediction which reigns at this point.

## 7. Province A: examining other methods

Section 6 served to show how the yardstick for measuring prediction quality introduced in Section 5 can be used to determine which of two variants or prediction methods tends to predict better for the data of Province A. Several other simple prediction methods and their variants were introduced for illustration in Table 1. The present section will examine how these would have performed had they been used.

### 7.1. Holt smoothing A[16]

For each of the 43 years 1955–1997 the predictions ($P(s)$, $s = 1, 2, \ldots, 10$) were computed. From these the relative differences $d(s)$ were obtained. The summary statistics $\bar{d}$, SE around $\bar{d}$ and SE around 0 are in Table 6.

These measures of performance are very similar to those obtained by linear regression with $n = 10$ as shown in Table 5. It follows that, in Province A, the Holt smoothing (A) is as good or slightly better as the extrapolation of a linear OLS fit with 10 data points. Both outperform the extrapolation of a quadratic fit.

### 7.2. Holt smoothing B[17]

The results for $s = 1$–10 for Holt smoothing B are in Table 7. Except at $s = 3$ they are not as good as those for Holt smoothing A.

### 7.3. Fitting a Hoerl function

The function $\alpha e^{\beta(t-\delta)}(t-\delta)^{\gamma}$ was fitted to $n = 5, 6, \ldots, 20$ data points by minimizing the sum of absolute deviations. How the $\bar{d}$ depends on $n$ and $s$ is shown in Fig. 16.

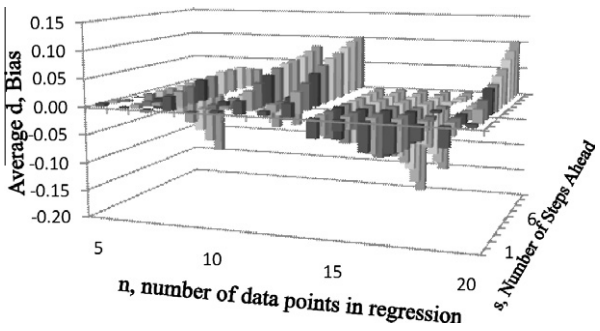


**Fig. 16.** Hoerl prediction: bias as a function of $s$ and $n$.

The shape of the bias function lacks regularity. It depends heavily on how well the chosen function fits the data and may be influenced by how well the Excel 'SOLVER' determines the minimum. Bias is minimal when $n = 7, 8$ or 9 and when $n = 18, 19$. How the two standard errors depend on $s$ and $n$ is shown in Figs. 17 and 18.

Considering the size of bias and of standard error, the best predictions for all $s$ are when $n = 19$. The measures of performance at $n = 19$ are in Table 8. They are significantly better than those of the linear prediction and the Holt smoothing A.

The measures of performance of the Hoerl function when $n = 10$ is in Table 9. These too are better than those of Holt smoothing A. Thus, for Province A, even when only 10 points were available for fitting the Hoerl function, the predictions were, on the average, better than by the other methods examined so far.

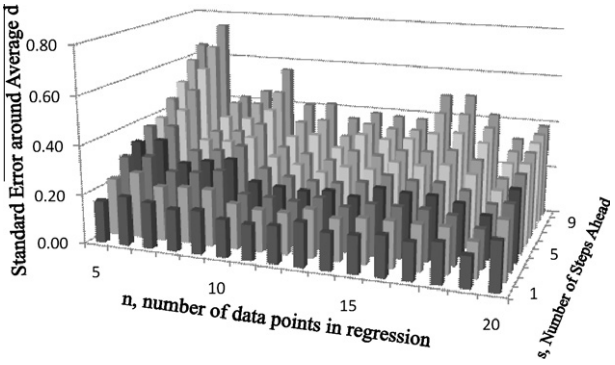


**Fig. 17.** Hoerl prediction: standard error around $\bar{d}$ as a function of $s$ and $n$.

[16] For description see Footnotes 9 and 10.
[17] See Footnote 12.

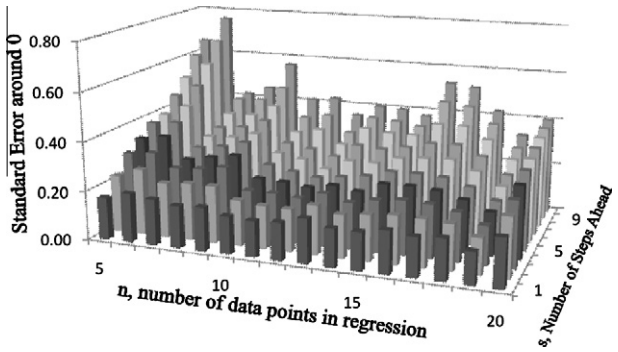


**Fig. 18.** Hoerl prediction: standard error around 0 as a function of $s$ and $n$.

### 7.4. The Oppe–Koornstra model

The Oppe–Koornstra model consists of two equations. The first equation traces the evolution of risk over time [$R_t = \exp(\alpha t + \beta)$ in which $R_t$ is the risk at time $t$, with $\alpha$ and $\beta$ as parameters to be estimated from data]. The second equations traces the evolution of exposure over time [$V_t = V_m/\{1 + \exp - [(a(t - c) + b]\}$ in which $V_t$ represents the vehicle-kilometres of travel ($V_{kmt}$) at time $t$, $V_m$ is the maximum (saturation) $V_{kmt}$ and a, b and c are parameters estimated from data]. The number of accidents is the product $R_t \times V_t$.

The $V_{kmt}$ information for Province A is available from the period 1960–2007. For each year between 1966 and 1998 the parameters $\alpha$, $\beta$, $a$, $b$, $c$ and $V_m$ were estimated using all the data from 1960 till that year. Using these parameters $R_t$ and $V_t$ were predicted for $s = 1, 2, \ldots, 10$ steps ahead and the relative differences [$d(s)$] calculated. The usual measures of performance are in Table 10. As is obvious the biases (the $\bar{d}$ and all the $d$'s that make it up) are negative and large. Had the Oppe–Koornstra model been applied to Province A it would have predicted more fatalities than what actually occurred. However, even if the bias is disregarded, the standard error around $\bar{d}$ is larger than when prediction is based on fitting a Hoerl function.

### 7.5. Summary for Province A

The performance of several methods and their variants for predicting the number of fatalities in Province A up to ten steps ahead was examined. It seems that of the many options compared, fitting a Hoerl function with $n = 10$ and $n = 19$ would have performed best. This conclusion has to be qualified in several ways. First, there are many other prediction methods and variants that could be tried. Thus, e.g., with Holt smoothing one could take the last $n$ $u$'s (slopes), extrapolate them in some suitable way, and use the sum of extrapolated $u$'s in the prediction. Or, e.g., one could use the Oppe–Koornstra model so that the fitted Risk and Exposure functions are based on the most recent $n$ data points, rather than all data points. Second, the method that, had it been used, would have performed best in Province A, may not be best in another jurisdiction. To draw more general conclusions, similar analyses need to be performed on data from many other jurisdictions.

**Table 8**
Measures of performance for Hoerl function $\{\alpha \exp[\beta(t-\delta)](t-\delta)^{\gamma}\}$ when $n = 19$.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | −0.03 | −0.02 | −0.02 | −0.02 | −0.02 | −0.02 | −0.03 | −0.03 | −0.03 | −0.03 |
| Standard error around $\bar{d}$ | 0.13 | 0.14 | 0.16 | 0.18 | 0.19 | 0.22 | 0.24 | 0.27 | 0.29 | 0.32 |
| Standard error around 0 | 0.13 | 0.15 | 0.16 | 0.18 | 0.19 | 0.22 | 0.24 | 0.27 | 0.30 | 0.33 |

**Table 9**
Measures of performance for Hoerl function $\{\alpha \exp[\beta(t-\delta)](t-\delta)^{\gamma}\}$ when $n = 10$.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | 0.00 | 0.00 | 0.00 | 0.00 | −0.01 | −0.02 | −0.04 | −0.05 | −0.07 | −0.08 |
| Standard error around $\bar{d}$ | 0.15 | 0.19 | 0.19 | 0.22 | 0.24 | 0.27 | 0.28 | 0.31 | 0.35 | 0.40 |
| Standard error around 0 | 0.15 | 0.19 | 0.19 | 0.22 | 0.25 | 0.27 | 0.29 | 0.31 | 0.35 | 0.41 |

**Table 10**
Measures of performance for the Oppe–Koonstra model.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | −0.30 | −0.34 | −0.39 | −0.43 | −0.47 | −0.51 | −0.56 | −0.59 | −0.64 | −0.69 |
| Standard error around $\bar{d}$ | 0.28 | 0.30 | 0.31 | 0.31 | 0.30 | 0.31 | 0.33 | 0.35 | 0.36 | 0.37 |
| Standard error around 0 | 0.42 | 0.46 | 0.50 | 0.53 | 0.56 | 0.59 | 0.65 | 0.69 | 0.74 | 0.78 |

**Table 11**
Predictions for 1986–1995.

| Year | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|---|---|---|---|---|---|---|---|---|---|---|
| Prediction | 216.6 | 210.0 | 203.4 | 196.6 | 189.8 | 183.0 | 176.2 | 169.4 | 162.7 | 156.1 |

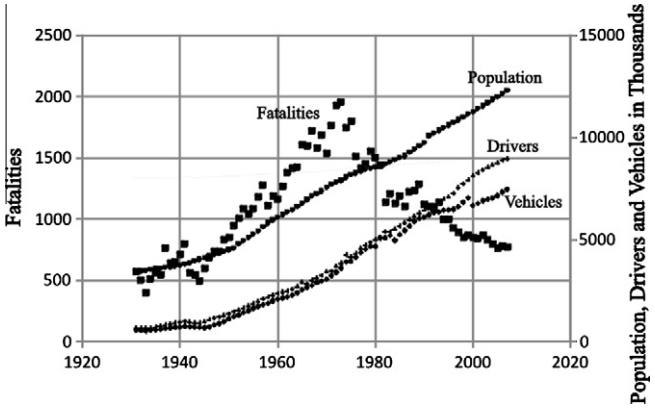


**Fig. 19.** Time series of fatalities, population, etc. in Province B.

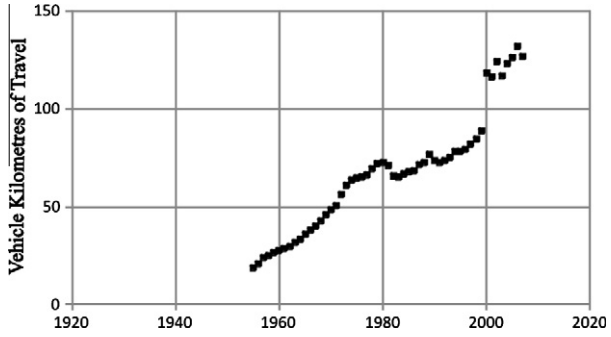


**Fig. 20.** Amount of travel in Province B.

To round out this section recall the story line around which the narration revolved. We imagined that we are at the end of 1985 and wish to predict for the years 1986–1995. It now seems that the Hoerl function with $n = 19$ would be sensible choice of method and variant. The parameter estimates based on the 1967–1985 data are $\alpha = 0.3438$, $\beta = -0.0982$, $\gamma = 2.822$ and $\delta = 1944.7$. Using these parameters the predictions are in Table 11.

Based on Table 8 the average bias is −0.02 to −0.03. It would therefore seem appropriate to reduce the prediction by 2–3%. With this the unbiased prediction, e.g., for 1990 ($s = 5$) is $189.8 \times 0.98 = 186.0$. Thus, at the end of 1985 one should predict that if the prevailing trends will continue and no substantially new policies and measures are implemented, in 1990 there will be in Province A about 186 fatalities. Also based on Table 8 the standard error is $\pm 0.19 \times 186 = \pm 35.3$ fatalities. Using the usual rule of thumb of ±2 standard deviations, the number of fatalities is almost certain to be in the 116–256 range.

## 8. Province B

The main time series information for Province B is in Figs. 19 and 20. While fatalities reached a peak and later declined, population, licensed drivers, and registered vehicles continued to increase over time. Till 1999 the $V_{kmt}$ information (in Fig. 20) was based on fuel sales data and after 2000 it is based on the Annual Canadian Vehicle Survey by Statistics Canada. Because of the uncertainties surrounding this data the number of Licensed Drivers is used as a proxy for Exposure.

### 8.1. Linear prediction

The performance of the linear OLS predictor on the Province B data is best for $n$ between 8 and 13 as shown in Table 12.

### 8.2. Quadratic prediction

For this method the quality of predictions improves with $n$. The measures of performance for $n$ from 15 to 19 are in Table 13. For $s = 1$ and 2 the quadratic predictor has a slight edge over the linear one but is significantly worse for larger $s$.

### 8.3. Holt smoothing A

It appears that the extrapolation of a linear fit to $n$ between 8 and 13 is about as good as the Holt Smoothing 'A'.

**Table 12**
Measures of performance for the linear OLS predictor.

| $n$ | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$, bias | | | | | | | | | | |
| 8 | −0.01 | −0.02 | −0.03 | −0.03 | −0.03 | −0.03 | −0.03 | −0.03 | −0.04 | −0.04 |
| 9 | −0.01 | −0.02 | −0.03 | −0.03 | −0.03 | −0.02 | −0.03 | −0.03 | −0.04 | −0.04 |
| 10 | −0.02 | −0.02 | −0.03 | −0.02 | −0.02 | −0.02 | −0.03 | −0.03 | −0.04 | −0.04 |
| 11 | −0.02 | −0.02 | −0.02 | −0.02 | −0.02 | −0.02 | −0.03 | −0.03 | −0.04 | −0.05 |
| 12 | −0.02 | −0.02 | −0.01 | −0.01 | −0.02 | −0.02 | −0.03 | −0.04 | −0.04 | −0.05 |
| 13 | −0.01 | −0.01 | −0.01 | −0.01 | −0.02 | −0.02 | −0.03 | −0.04 | −0.05 | −0.06 |
| SE around $\bar{d}$ | | | | | | | | | | |
| 8 | 0.13 | 0.17 | 0.22 | 0.24 | 0.26 | 0.28 | 0.31 | 0.35 | 0.39 | 0.42 |
| 9 | 0.13 | 0.18 | 0.21 | 0.23 | 0.25 | 0.26 | 0.30 | 0.33 | 0.37 | 0.40 |
| 10 | 0.14 | 0.17 | 0.20 | 0.22 | 0.23 | 0.26 | 0.29 | 0.32 | 0.36 | 0.39 |
| 11 | 0.14 | 0.17 | 0.19 | 0.20 | 0.23 | 0.26 | 0.29 | 0.32 | 0.35 | 0.38 |
| 12 | 0.13 | 0.16 | 0.18 | 0.20 | 0.23 | 0.25 | 0.29 | 0.32 | 0.35 | 0.38 |
| 13 | 0.13 | 0.15 | 0.17 | 0.20 | 0.23 | 0.26 | 0.29 | 0.32 | 0.35 | 0.39 |
| SE around 0 | | | | | | | | | | |
| 8 | 0.13 | 0.18 | 0.22 | 0.24 | 0.26 | 0.28 | 0.31 | 0.35 | 0.39 | 0.42 |
| 9 | 0.14 | 0.18 | 0.21 | 0.23 | 0.25 | 0.27 | 0.30 | 0.34 | 0.37 | 0.40 |
| 10 | 0.14 | 0.18 | 0.21 | 0.22 | 0.23 | 0.26 | 0.30 | 0.33 | 0.36 | 0.39 |
| 11 | 0.14 | 0.17 | 0.19 | 0.20 | 0.23 | 0.26 | 0.29 | 0.32 | 0.36 | 0.39 |
| 12 | 0.14 | 0.16 | 0.18 | 0.20 | 0.23 | 0.25 | 0.29 | 0.32 | 0.36 | 0.39 |
| 13 | 0.13 | 0.15 | 0.18 | 0.20 | 0.23 | 0.26 | 0.29 | 0.32 | 0.36 | 0.39 |

**Table 13**
Measures of performance for the quadratic OLS predictor.

| $n$ | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$, bias | | | | | | | | | | |
| 15 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | −0.01 | −0.01 |
| 16 | 0.00 | 0.00 | 0.00 | 0.00 | −0.01 | −0.01 | −0.01 | −0.02 | −0.03 | −0.03 |
| 17 | 0.00 | −0.01 | −0.01 | −0.01 | −0.02 | −0.02 | −0.03 | −0.04 | −0.05 | −0.05 |
| 18 | −0.01 | −0.01 | −0.02 | −0.02 | −0.03 | −0.03 | −0.04 | −0.05 | −0.05 | −0.06 |
| 19 | −0.01 | −0.01 | −0.02 | −0.02 | −0.03 | −0.03 | −0.04 | −0.05 | −0.05 | −0.06 |
| 20 | −0.01 | −0.01 | −0.02 | −0.03 | −0.03 | −0.03 | −0.04 | −0.04 | −0.05 | −0.05 |
| SE around $\bar{d}$ | | | | | | | | | | |
| 15 | 0.12 | 0.17 | 0.22 | 0.28 | 0.35 | 0.41 | 0.50 | 0.59 | 0.68 | 0.78 |
| 16 | 0.12 | 0.16 | 0.22 | 0.27 | 0.33 | 0.40 | 0.48 | 0.56 | 0.65 | 0.75 |
| 17 | 0.12 | 0.17 | 0.22 | 0.27 | 0.33 | 0.39 | 0.47 | 0.55 | 0.64 | 0.73 |
| 18 | 0.12 | 0.17 | 0.22 | 0.27 | 0.32 | 0.38 | 0.46 | 0.54 | 0.63 | 0.72 |
| 19 | 0.13 | 0.17 | 0.22 | 0.26 | 0.32 | 0.38 | 0.45 | 0.53 | 0.62 | 0.71 |
| 20 | 0.13 | 0.17 | 0.21 | 0.26 | 0.32 | 0.38 | 0.45 | 0.52 | 0.61 | 0.70 |
| SE around 0 | | | | | | | | | | |
| 15 | 0.12 | 0.17 | 0.22 | 0.28 | 0.35 | 0.41 | 0.50 | 0.59 | 0.68 | 0.78 |
| 16 | 0.12 | 0.16 | 0.22 | 0.27 | 0.33 | 0.40 | 0.48 | 0.56 | 0.66 | 0.75 |
| 17 | 0.12 | 0.17 | 0.22 | 0.27 | 0.33 | 0.39 | 0.47 | 0.55 | 0.64 | 0.74 |
| 18 | 0.12 | 0.17 | 0.22 | 0.27 | 0.32 | 0.38 | 0.46 | 0.54 | 0.63 | 0.72 |
| 19 | 0.13 | 0.17 | 0.22 | 0.26 | 0.32 | 0.38 | 0.46 | 0.53 | 0.62 | 0.71 |
| 20 | 0.13 | 0.17 | 0.22 | 0.26 | 0.32 | 0.38 | 0.45 | 0.52 | 0.61 | 0.70 |

**Table 14**
Measures of performance for Holt smoothing A.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | 0.00 | −0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | −0.01 | −0.03 | −0.04 |
| Standard error around $\bar{d}$ | 0.11 | 0.15 | 0.20 | 0.21 | 0.21 | 0.23 | 0.28 | 0.31 | 0.36 | 0.40 |
| Standard error around 0 | 0.11 | 0.15 | 0.20 | 0.21 | 0.22 | 0.23 | 0.28 | 0.31 | 0.36 | 0.40 |

**Table 15**
Measures of performance for the Hoerl function.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.07 | 0.08 | 0.09 |
| Standard error around $\bar{d}$ | 0.16 | 0.18 | 0.19 | 0.21 | 0.24 | 0.26 | 0.29 | 0.31 | 0.34 | 0.36 |
| Standard error around 0 | 0.16 | 0.18 | 0.19 | 0.22 | 0.24 | 0.27 | 0.29 | 0.32 | 0.35 | 0.37 |

### 8.4. The Hoerl function

For the Province B time series of fatalities the Hoerl Function performs best when $n$ is about 12. The measures of performance at this point are in Table 15.

### 8.5. The Oppe–Koornstra method

To apply this method one has to compute risk which, in the Oppe–Koornstra formulation is the ratio of fatalities per vehicle-kilometres of travel. Inasmuch as for Province B the information about $V_{kmt}$ is sketchy as shown in Fig. 20, the number of licensed drivers will be the proxy for exposure.[18] Even though the predictions of risk and exposure are quite good, their product does not yield satisfactory predictions as is obvious from comparing the measures of performance in Table 16 to these in Table 12 or Table 14.

For Province B the count of fatalities begins in 1931. To produce Table 16 the parameters of the risk and exposure function were estimated using data from 1931 till the year before the prediction for $s = 1$. This is in line with the spirit of the Oppe–Koornstra thinking which regards the regularities in the evolution of risk and exposure as representing some deeper theoretical principles. It is possible the predictions would be better if the parameter estimates were based on the $n$ most recent data points rather than on all available data.

### 8.6. Summary for Province B

The performance of several methods and their variants for predicting the number of fatalities in Province B for up to ten steps (years) ahead was examined. It seems that the linear predictor with $n$ around 10, Holt smoothing A, and the Hoerl function predictor perform similarly. The quadratic predictor was better than the

[18] $V_{kmt}$ and the number of licensed drivers are usually closely correlated. It is possible that the number of kilometres of travel per licensed driver changes gradually over time. However, in the period 2000–2007 for which reliable $V_{kmt}$ data are available there is no evidence of any such trend.

**Table 16**
Measures of performance for the Oppe–Koornstra method.

| | Steps ahead | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $\bar{d}$ = average $d$ (bias) | −0.09 | −0.10 | −0.12 | −0.13 | −0.15 | −0.17 | −0.19 | −0.20 | −0.21 | −0.22 |
| Standard error around $\bar{d}$ | 0.31 | 0.33 | 0.34 | 0.36 | 0.37 | 0.38 | 0.39 | 0.41 | 0.42 | 0.44 |
| Standard error around 0 | 0.32 | 0.34 | 0.36 | 0.38 | 0.40 | 0.42 | 0.44 | 0.46 | 0.47 | 0.49 |

linear one for one or two steps ahead and significantly worse for predicting further into the future. The Oppe–Koornstra method has a relatively large average bias and standard error. It is possible that its performance would improve if the number of data points used in the curve fitting was limited.

## 9. Summary and discussion

Prediction is about potential outcomes: *what will happen if* and about *what would have happened if*. As there are many ways to predict one has to determine what method tends to predict best. To do so empirically one asks what method would have predicted best had it been applied in the past and assumes, inductively, that the same will be true in the future.

To say which of two methods is better, one has to have a yardstick for prediction quality. It is suggested that $\bar{d}$ (the average bias) and the standard error of $d$ be used for this purpose. The $\bar{d}$ tells by what proportion, on the average, the count of accidents would exceed the prediction if the method was applied year after year. The standard error (around $\bar{d}$ or around 0) measures the variability of the $d$'s in the usual manner. It is shown how these measures of prediction quality allow one to determine which of two methods should be preferred in what circumstance.

The approach was applied to two data sets: The time series of motor vehicle accident fatalities in Province A and in Province B. On the basis of this analysis one may draw tentative conclusions for these jurisdictions and the methods tested; one can say what method seems preferable, what is the average size of bias that needs to be corrected, and how accurate is the prediction likely to be.

Broader conclusions will emerge once many additional methods of prediction are applied to data from many other jurisdictions and pertaining to other circumstances. As there are many jurisdictions that have long time series of road safety related data, research along these lines is both feasible and attractive. The hope is that the approach outlined here will prove attractive to others and that the cumulative results of research on prediction will make for better safety evaluation and improved setting of safety targets.

## Acknowledgements

The help of Transport Canada, Road Safety is gratefully acknowledged. Kwei Quaye of the Saskatchewan Government Insurance, Leo Tasca and Chris Janusz of the Ontario Ministry of Transport provided important data.

## References

Broughton, J., Allsop, R.E., Lynam, D.A., McMahon, C.M., 2000. The numerical context for setting national casualty reduction targets. TRL Report 382, TRL Limited, Wokingham.
Broughton, J., 2006. Monitoring progress towards the GB casualty reduction target. In: Proceedings, European Transport Conference, Strasbourg, France.
Hauer, E., Ng, J.N.C., Papaioannou, P., 1991. Prediction in road safety studies: an empirical inquiry. Accident Analysis and Prevention 23 (6), 595–607.
Hauer, E., 1991. Comparison groups in road safety studies: an analysis. Accident Analysis and Prevention 23 (6), 609–622.
Hauer, E., 1997. Observational Before–After Studies in Road Safety. Pergamon.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22 (4), 679–688.
Makridakis, S., Hibon, M., 2000. The M3-competition: results, conclusions, and implications. International Journal of Forecasting 16, 451–476.
Oppe, S., Koornstra, M.J., 1990. A mathematical theory for related long term developments of road traffic and safety. In: Koshi, M. (Ed.), Transportation and Traffic Theory. Elsevier, New York, pp. 113–132.
Oppe, S., 1991. Development of traffic and traffic safety: global trends and incidental fluctuations. Accident Analysis and Prevention 23 (5), 413–422.
Quaye, K.E., 1992. Forecasting models in road safety studies. PhD Dissertation, Department of Civil Engineering, University of Toronto.
Quaye, K.E., Hauer, E., 1993. The use of forecasting models in the evaluation of safety interventions: a theoretical inquiry. In: Daganzo, C.F. (Ed.), Transportation and Traffic Theory. Elsevier Science Publishers, pp. 313–332.
Quaye, K.E., Hauer, E., 1994. Assessing forecasting methods used in before after studies. Paper presented at the 72nd Annual Meeting of the Transportation Research Board, Washington.
Rubin, D.B., 2005. Causal inference using potential outcomes: design, modeling, decisions. Journal of the American Statistical Association 100 (469), 322–331.