The Magnitude Of The Regression To The Mean Effect In Traffic Crashes

Ellen De Pauw
Hasselt University – Transportation Research Institute
Wetenschapspark 5, 3590 Diepenbeek, Belgium
+3211269157
Ellen.Depauw@Uhasselt.Be

Stijn Daniels
Hasselt University – Transportation Research Institute
Wetenschapspark 5, 3590 Diepenbeek, Belgium
+3211269156
Stijn.Daniels@Uhasselt.Be

Tom Brijs
Hasselt University – Transportation Research Institute
Wetenschapspark 5, 3590 Diepenbeek, Belgium
+3211269155
Tom.Brijs@Uhasselt.Be

Hermans Elke
Hasselt University – Transportation Research Institute
Wetenschapspark 5, 3590 Diepenbeek, Belgium
+3211269141
Elke.Hermans@Uhasselt.Be

Wets Geert
Hasselt University – Transportation Research Institute
Wetenschapspark 5, 3590 Diepenbeek, Belgium
+3211269158
Geert.Wets@Uhasselt.Be

Abstract
To evaluate a traffic safety measure often the crash rates after the implementation is compared with the crash rates before the measure. One of the most challenging problems using this study design, are the confounding variables, whose effect can lead to false estimations about the effectiveness of a traffic safety measure. Regression to the mean is often defined as one of the most important confounding variables. Past studies showed the existence of RTM in traffic crashes, however some lack of clarity exist concerning this phenomenon. The magnitude of regression to the mean was examined through a sub dataset of the black spot program in Flanders. This research group consisted of 253 intersections that knew a higher than average number of crashes during 1997-1999. This number of crashes was compared with the next three years, during which no traffic safety measures were applied, to see whether regression to the mean occurred. This comparison
from before to after, showed a high decrease in number of crashes, both for crashes with all injuries, as for crashes with serious injuries and fatalities. To count which part of this decrease could be accounted to other confounding variables, such as the trend effect and chance, a comparison group was used. This comparison group consisted of all crashes at intersections with a regional road, since all the sites from the research group are located on those intersections. Even when trend was controlled, a high decrease in number of injury crashes was found. This decrease can subsequently be attributed to the regression to the mean. For crashes with serious injuries and fatalities, the magnitude of the regression to the mean was lower, however still present.

**Introduction**

To handle an effective traffic safety policy, there is an undisputed need to evaluate the effects of traffic safety measures. Before-and-after study (B&A) are stated as the most appropriate method to execute this (Elvik, 2002; Shinar, 2007). Different B&A methods are available, which mainly differ in the extent they control for possible confounding variables. A confounding variable is defined as any exogenous variable affecting the number of crashes or injuries whose effects, if not estimated, can be mixed up with the measure being evaluated (Elvik, 2002). Variables that are regarded as potentially confounding in observational B&A studies of road safety measures are regression to the mean, long term trend affecting the number of crashes or injured road users, general changes in the number of crashes, changes in traffic volumes and any other specific events introduced at the same time as the road safety measure (Elvik, 2002; Hauer, 1997; Hirst, Mountain & Maher, 2004).

Regression to the mean (RTM) is defined as one of the most important confounding variables (Hauer, 1997), and this phenomenon was already stated quite some years ago (Tamburri, Hammer, Glennon & Lew, 1968). "Regression-to-the-mean denotes the tendency for an abnormally high number of accidents to return to values closer to the long term mean; conversely abnormally low numbers of accidents tend to be succeeded by higher numbers. Regression-to-the-mean occurs as a result of random fluctuation in the recorded number of accidents around the long-term expected number of accidents" (Elvik & Vaa, 2004). As often the decision to implement a measure is based on a high number of crashes, it is conceivable that the number of crashes will decrease, irrespective of the measure. In those cases RTM will lead to an overestimation of the treatment effectiveness when not appropriately taken into account. Based on a large number of diverse datasets, the substantial effect of the RTM was demonstrated (for example Hauer, 1997; Hauer & Persaud, 1983).

**Background**

Many B&A studies controlled for RTM when examining the effectiveness of a traffic safety research. One of the first articles is that of Hauer and Persaud (1983), who stated that simple B&A studies tend to overestimate the safety effectiveness of a measures, due to the RTM phenomenon. To prove this, they used several real-world data sets. Elvik (2002) compared studies that did not controlled for RTM, with studies that did controlled. From this it seemed that the magnitude of RTM was estimated to be 20-30% in three of seven cases, but was of a negligible magnitude in the other four. Calculating the effects that are attributable to the measures, when RTM and other general changes where controlled, ranged from 6 to 52% reduction of the number of crashes. Harwood et al. (2007) applied three B&A methods to evaluate the safety-effects of the installation of added left-turn lanes, added right-turn lanes and the extension of the length of existing left- or right-turn lanes. The Empirical Bayes method, that controlled for RTM, resulted in lower effectiveness estimates, compared to the yoked comparison and the comparison group approach, that did not controlled for RTM. The authors attributed this difference to the RTM phenomenon. In a study that analyzed the effectiveness of fixed and mobile speed and red light cameras
(Gains, Nordstrom, Heydecker & Shrewsbury, 2005), the authors also examined the RTM effects on a subset of 216 cameras. Therefore they applied the empirical Bayes method with an external crash model. The analysis showed that RTM only had a modest effect on all injury crashes, that is an average fall of 7%, but had an appreciable effect on fatal and serious injury crashes of 35%. Compared to the total decrease, the RTM represented one quarter of the observed fall in injury crashes, and three fifths of the observed fall in more severe crashes. A meta-analysis that examined the effectiveness of red light cameras (Erke, 2008) showed that studies that controlled for RTM and spillover effects, yield less favorable results, than results from studies that did not controlled these confounding variables. Also here it is the case that RLC are installed at intersections with an exceptionally high number of crashes, that leads possible to higher RTM. The study concluded with the assumption that a failure to control for RTM can lead to an overestimation of the effect of RLC (Erke, 2008).

Next to the effect of RTM in the evaluation of traffic safety measures, a couple of researchers also examined explicitly the RTM phenomenon. Sharma & Datta (2007) examined the RTM effects using four B&A evaluation methods (B&A, B&A with comparison group, 2 EB methods) to evaluate the effectiveness of low-cost safety improvements. From this it was found that studies that did not controlled for RTM, produce similar results to those that did controlled for RTM, when three to five years of crash rates were used. The authors concluded that the RTM effect becomes insignificant when three or more years of crash data are used in the evaluation of high crash locations. On a similar way, Persaud & Lyon (2007) examined the RTM effect at 1669 rural, 4-legged stop controlled intersections in California, from 1994-1996 to 1997-1999. However here no measure was taken and the sites remained unaltered during those periods. To measure the RTM effect, the number of crashes for 1997-1999 were estimated, on the basis of crash rates from 1994-1996, using first a comparison ratio of the quotient of the mean number of crashes of the before and after period. Secondly the empirical Bayes method method was applied, that controlled for RTM. This empirical Bayes method showed a result more closer to the observed number. Next to this, the authors searched for studies that estimated the effectiveness of a measure trough the empirical Bayes method. To see to what extent RTM was present in these studies, they applied a naïve B&A method. In terms of % reduction, this resulted in a relatively small difference between the naïve and the empirical Bayes estimates. However the safety effect in term of the actual reduction in crash frequency, which is stated as a more important factor, showed substantial differences between the empirical Bayes and the naïve estimates. The RTM phenomenon was stated as the main reason for this difference (Persaud & Lyon, 2007).

From those studies the existence and importance of the RTM phenomenon is clearly shown. However, those results are not always consistent, and vary along different elements, such as the years that are included in the before and after period, the investigated measure, the crash type included... Uncertainty about the RTM effect raises. Skeptics belief that when many years of pre-treatment data are used to select entities for treatment or in an evaluation, and those entities have a high number of crashes, there will be little or no RTM (Persaud & Lyon, 2007). However it is difficult to establish how many years of pre-treatment are required, or how high crash counts need to be (Persaud & Lyon, 2007).

Data
To measure the RTM effect the data set from the black spot program in Flanders was used. To increase traffic safety and reach the goals of a 50% decrease in crashes with serious injuries and fatalities in 2010 compared to 1999 (Ministerie van de Vlaamse Gemeenschap, 2001), the government decided in 2002 to, next to general education and enforcement, adapt the infrastructure of the most dangerous traffic points in Flanders. A location was selected as ‘dangerous’ when during the years 1997-1999 at least 3 injury crashes occurred, and a total severity score of 15 was reached. This score was calculated through the assignment of a weight of 1 to every slightly injured (defined as every person that got
injured in a traffic crash, but cannot be defined as severely or deadly injured), 3 to seriously injured (every person that got involved in a traffic crash and a hospitalization of more than 24 hours was necessary) and 5 to deadly injured (every person that as a consequence of a traffic crash died on the spot, or within 30 days after the crash). A total of 1014 black spots were selected, of which the 800 most dangerous were planned to be adapted. In the light of the theory of the RTM phenomenon, it can be expected that those sites with this (extreme) number of crashes, would return to a more average number in the years after, even without the implementation of a traffic safety measure. This dataset was perfectly suitable to calculate the RTM, as the adaptations of the locations started in 2003. During the period that was used to select the black spots and the starting point of the first adaptations, also called as the lag period (Maher & Mountain, 2009), no traffic safety measures were taken on these locations. From this period (2000-2002) it could be examined whether RTM occurred.

In this study, the research group encompasses a subset of the most dangerous spots. To select those sites not the priority score was used, as this also counts for the severity of the crashes, but the focus was on the number of crashes. Therefore the 300 sites with the highest number of crashes were included. The majority of the 300 sites are intersections, only a small number of the sites are road sections. Because a different radius was used to select the number of crashes for road sections compared to intersections, only intersections were selected. This resulted in a research group of 253 sites. These sites are all situated at intersections at regional roads.

To select a comparable comparison group, a yoked comparison method was used. For each of the black spots, a location in the comparison group was selected, so that there is a one-to-one correspondence between each site in the comparison and the research group. Therefore sites were selected that were in the vicinity of the research group, by searching on the same road for the next intersection with the same kind of crossing roads. To make sure the comparison group does not include sites with high crash rates, the 1014 sites that were defined as black spot were excluded from the comparison group. To select this comparison group the spatial analysis program ArcGIS, version 9.3 was used.

To select the crashes that occurred at the research and comparison locations also ArcGIS was used. All crashes in a radius of 50 meters around the location were selected. As can be seen from figure 1, the selected comparison group shows an increase in number of crashes. This increase was significant for both all injury crashes (F= 13,64; p<0,001) and crashes with serious injuries and fatalities (F= 7,46; p= 0,007). As the common crash trend in Flanders and Belgium shows a decrease, this is remarkable. A possible explanation could be that the 1014 sites that were defined as most dangerous, were excluded from the comparison group. This could possibly have led to a selection of comparison locations with an unusual low number of crashes. It could even be said that this comparison group shows the other direction of RTM: an (extreme) low number of crashes returns to a higher, more average number of crashes. Such a comparison group gives a false estimation of the trend and would lead to unreliable results. Therefore another comparison group was chosen. Since all the sites from the research group are located at intersections with a regional road, all crashes that occurred in Flanders at intersections with a regional road were selected. This selection shows a decrease in number of crashes over the years (see figure 1). From figures 2 and 3 can be seen that this is the case for both all injury crashes (all crashes with at least a severely injured), and the more severe crashes (only crashes with severe injuries and fatalities). Subsequently this is a better reflection of the general crash trend in Flanders and Belgium. The disadvantage of this selection is that the research sites are included in the comparison group. However, the most important issue is that the comparison group gives a good idea of the general trend.
FIGURE 1: Percent difference in number of crashes between research group, the comparison group selected through the yoked method, and the comparison group including all injury crashes at intersections at regional roads.

FIGURE 2: Number of injury crashes in research and comparison group.
FIGURE 3: Number of severe crashes in research and comparison group

To examine whether the comparison group is comparable to the research group, the odds ratio (OR) for the crash rates during the before period can be used (Fleiss, 1981). However here the OR is applied on the whole study period, as it could be expected that both groups are similar over those period, since no measure is executed.

\[
\text{OR} = \frac{R_t}{R_{t-1}} \times \frac{C_{t-1}}{C_t}
\]

\(R_t\) = number of crashes in research group in year \(t\)
\(R_{t-1}\) = number of crashes in research group in year \(t-1\)
\(C_t\) = total number of crashes in comparison group in year \(t\)
\(C_{t-1}\) = total number of crashes in comparison group in year \(t-1\)

An OR near to 1, indicates that the comparison group is comparable to the research location. Maximum standard deviation should not be higher than 0,20. The OR here is calculated for the total number of crashes per year. The results in table 1 show that for most of the years the odds ratio is near to 1, and the standard deviation is lower than 0,20. We can conclude that this is a fairly good comparison group. Furthermore it can be expected that given fact that the comparison group encompasses crashes from all sorts of intersections at regional roads, this is a good reflection of the characteristics of the sites in the research group.

**TABLE 1: Odds ratio and standard deviation per year, both for injury and more severe crashes**

<table>
<thead>
<tr>
<th></th>
<th>Injury crashes OR (s)</th>
<th>Severe crashes OR (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>97-98</td>
<td>0,95 (0,04)</td>
<td>0,97 (0,10)</td>
</tr>
<tr>
<td>98-99</td>
<td>1,13 (0,04)</td>
<td>0,89 (0,10)</td>
</tr>
<tr>
<td>99-00</td>
<td>0,76 (0,04)</td>
<td>0,94 (0,10)</td>
</tr>
<tr>
<td>00-01</td>
<td>0,90 (0,05)</td>
<td>0,95 (0,11)</td>
</tr>
</tbody>
</table>
Study Design
To examine the magnitude of the RTM phenomenon, it was examined to what extent the number of crashes changed, during the period after the period that was used to select the sites. The before period for those analyses is the period that is used to select the black spots, which is 1997-1999. The after period encompasses the next three years (2000-2002), during which no measures were executed yet, also called as the ‘lag-period’. At first the naïve B&A method was applied. This method shows how the number of crashes has evolved over the years, without any control for confounding variables. Secondly the B&A method with a comparison group is used, which calculates the change in number of crashes, taking trend into account. This is important as the overall crash trend will change over the years, dependent on for example new safety technologies, legislation, education… Those number of crashes, that is the recorded number after controlled and not controlled for trend, will be compared to the recorded number of crashes from the before period. As no traffic safety measure was applied, it could be expected that controlling the number of crashes for trend, would lead to the same results as recorded during the before period. When however a difference in number of crashes will be found between those two periods, this difference can be ascribed to RTM.

Both methods were applied on injuries crashes and on severe crashes. Additionally the analyses were repeatedly applied on all three years of before (’97-’99) and after crash data (’00-’02), two years of before (’98-’99) and after (’00-’01) data and one year of before (’99) and after (’00) data. From this it can be examined whether there is a difference in the magnitude of RTM, according to the number of years of data included.

Method
To calculate the magnitude of the RTM effect, the number of crashes from the after period are compared to the number of crashes from the before period. This comparison can be expressed by an index, similar to the index of effectiveness (Hauer, 1997), however slightly different as no measure was applied. The proportion of the number of crashes after to before, can be expressed by a relative count that indicates how much of the crashes remains during the after period:

\[
\text{Index of difference} = \frac{\text{counted number of crashes during the after period}}{\text{counted number of crashes during the before period}}
\] (2)

When the index is equal to 1, this means that there was no difference in number of crashes. An index lower than 1 indicates that the number of crashes decreased. An index higher than 1, shows an increase in number of crashes after, compared to before. Dependent on the method that is used, the counted number of crashes during the before period has to be controlled for confounding variables.

Naïve B&A
The naïve B&A method does not control for any confounding variable, and subsequently the denominator consists of the recorded number of crashes in the before period.

\[
\text{Index of difference} = \frac{L_{\text{after}}}{L_{\text{before}}}
\] (3)
With:
L<sub>after</sub> = number of crashes on location L during the after period
L<sub>before</sub> = number of crashes on location L during the before period

Reliability is measured through the 95% confidence interval. The index is an odds-ratio with lognormal distribution. Ln(eff) is the logarithm of the odds-ratio.

95% confidence interval (CI):
EFF, below limit = exp[ln(EFF) – 1.96 * s]
EFF, above limit = exp [ln(EFF) + 1.96 * s]

(4)

With the standard deviation (s) as the root of the variance (s²):

\[ s^2 = \frac{1}{L_{after}} + \frac{1}{L_{before}} \]  

(5)

Additional attention is needed when the number of crashes in the before or after period encompasses zero crashes. In those cases it is not possible to apply above equations. To make a calculation possible, these numbers were heightened with 0,5 (Elvik, 1997a; Fleiss, 1981). A possible disadvantage is that this can lead to slight deviant results, certainly when a lot of sites has to be heightened. However it still gives a realistic and good estimation.

**B&A with comparison group**

A second method is the B&A study with comparison group, which controls for trend effects. Applying control for trend through a comparison group, leads to next equation:

\[
\frac{L_{after}}{L_{before} \ \cdot \ \frac{C_{after}}{C_{before}}}
\]

Index of difference  

(6)

Which can be written as an odds ratio:

\[
\frac{L_{after}}{L_{before}} = \frac{L_{after}}{L_{before} \ \cdot \ \frac{C_{after}}{C_{before}}}
\]  

(7)

L<sub>after</sub> = number of crashes on location L during after period
L<sub>before</sub> = number of crashes on location L during before period
C<sub>after</sub> = number of crashes in comparison group during after period
C<sub>before</sub> = number of crashes in comparison group during before period

Reliability is measured as given by equation (3) and the variance:

\[ s^2 = \frac{1}{L_{after}} + \frac{1}{L_{before}} + \frac{1}{C_{after}} + \frac{1}{C_{before}} \]  

(8)
Meta-analysis
In addition to an individual analysis per location, a meta-analysis can be executed, which calculates the overall effect of all locations included per method. This combination leads to a result that is more statistically reliable (Fleiss, 1981). As shown above a measure on a certain place can be presented by equation (2) for the naïve method and (6) for the B&A method with comparison group, with accompanied confidence intervals and variances. To count the combined effect per method, every location gets an importance. This is inversely proportional to the variance. The higher the data of a site, the smaller the variance becomes, and the higher the importance. Sites with a high reliability, get the highest importance.

$$\frac{1}{w_i} = \frac{1}{\sigma^2_i}$$

(9)

Supposing that dispose of different places where a measure is executed, the weighted mean index of difference of the measure over all places is:

Overall index of difference = \[ \exp\left( \frac{\sum_{i=1}^{n} w_i \cdot \ln(EFF_i)}{\sum_{i=1}^{n} w_i} \right) \]

(10)

The estimation of a 95% confidence interval is

95% CI = \[ \left[ \frac{\sum_{i=1}^{n} w_i \cdot \ln(EFF_i)}{\sum_{i=1}^{n} w_i} \pm 1.96 \cdot \frac{1}{\sqrt{\sum_{i=1}^{n} w_i}} \right] \]

(11)

Not in all cases it is possible to estimate the overall effect. A possible problem that could appear here, is that the results of each site largely differ. To see this possible problem a funnel graph has to be drawn. In a funnel graph the indexes per location are set off against the weight of every location. When the overall estimation is allowed to be measured, the figure shows an upside down funnel. The top of the funnel is the site with the highest weight, and also with highest number of data, and consequently the most reliable one. When there is more than one top, or no top at all, the overall effect cannot be estimated.

Results
At first both methods were applied on every site, separately for injury crashes and more severe crashes, taking one to three years of B&A data into account. For every method, per years of included B&A data of injury and severe crashes, it was examined whether an overall estimation was possible. Therefore funnel plots were drawn, which all showed an upside down funnel. When only taking into account one year of B&A crash data with serious injuries and fatalities, the top was less clear, however still present. From this we can conclude that an overall estimation of the index of difference was possible. Those results are shown in table 2.

For the injury crashes the naïve B&A method clearly shows a significant index smaller than 1, which indicates a decrease in the number of crashes. On average a decrease of 25% is calculated, not fluctuating a lot when different numbers of years of crash data are included. These are quite high decreases as no specific measures were applied at these locations. As a consequence those decreases will have to be attributed to confounding variables, such as trend effects, RTM, and chance (Elvik, 2002). When trend is taken into account, using the B&A method with comparison group, the results show an index that is still significant lower than 1, but is higher compared to the naïve method. This means that the number of crashes during the after period are higher when trend is taken into account. Or,
stated differently, the decrease from before to after is lower when trend is taken into account. This means that trend accounts for a part of the recorded decrease in injury crashes. However even with this correction, still a high decrease was found. This difference can subsequently be attributed to the RTM effect.

For the severe crashes, the naïve B&A method also shows a decrease of 25% when three years of B&A data were taken into account. For two years of data the decrease is only 15%, for one year of B&A data, no significant difference in number of crashes between the before and after period is found. However it can be stated that the most reliable estimations come from the account of three years of B&A data. When the trend is taken into account, this results in an overall index of 0.93 for three year of B&A data, and 0.94 for two years. This is not significant, however the upper bound is close to 1. We can conclude that, however quite low, RTM is still present for severe injuries.

A possible explanation for this higher index for severe crashes compared to injury crashes, that resulted from the B&A method with comparison group, is that the decreasing trend effect in the comparison group is quite stronger for the severe crashes, compared to the injury crashes. The injury crashes show a mean decrease of 10%, but when only the severe crashes are taken into account, it is seen that these decline with about 18%. Another possible explanation is that the research group is selected on basis of the number of all crashes, and no distinction was made according to the severity of the crashes. It is possible that the research group especially exist of sites that during 1997-1999 had a high number of crashes with slightly injured, but with only an average number of severe crashes. These severe crashes will not show a RTM, as these were not higher than average during the before period.

The number of crashes for the after period, controlled for trend, can be estimated by the product of the number of crashes in the before period and those indexes. The results of this calculation, together with the recorded number of crashes from the before and after period, are shown in table 3, and displayed at figures 3 and 4. This table also displays the relative difference between the number of crashes in the before period and the number of crashes in the after period, both controlled and not controlled for trend. This shows that there is quite a difference between the number of crashes from before to after period, also when trend is taken into account. Especially for injury crashes high numbers are found. It can also be seen that there are no high difference in RTM when different number of years of B&A data are taken into account. Figures 3 and 4 also show for each of the years taken into account, the
number of crashes. However these number are given as a mean estimate per year, since a better comparison is possible between the different account for number of years. The vertical distance between the variables shows the mean difference in numbers of crashes per year. Taking into account the trend effect for injury crashes, this results in an estimation that is closer to the recorded number of crashes before. A part of the difference between the number of crashes before and after can be explained by the general trend. The remaining difference can subsequently be ascribed to the RTM effect. Figure 4 shows somewhat other results for more severe crashes. Here it can be seen that, taking trend into account, almost fully explains the contrast between the recorded number before and after. Only a small number is attributable to the RTM.

TABLE 3: Recorded number of crashes in the before period and the after period, both for no control for any confounding variable, and with control for trend effects

<table>
<thead>
<tr>
<th>Number of B&amp;A data taken into account</th>
<th>Recorded before</th>
<th>Recorded after</th>
<th>After with control for trend</th>
<th>Relative difference recorded after - before</th>
<th>Relative difference after controlled for trend - before</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury crashes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>1568</td>
<td>1152</td>
<td>1270</td>
<td>-36,11</td>
<td>-23,46</td>
</tr>
<tr>
<td>2 year</td>
<td>2940</td>
<td>2131</td>
<td>2352</td>
<td>-37,96</td>
<td>-25</td>
</tr>
<tr>
<td>3 year</td>
<td>4339</td>
<td>3153</td>
<td>3558</td>
<td>-37,62</td>
<td>-21,95</td>
</tr>
<tr>
<td>Severe crashes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 year</td>
<td>229</td>
<td>202</td>
<td>229</td>
<td>-13,37</td>
<td>0</td>
</tr>
<tr>
<td>2 year</td>
<td>474</td>
<td>373</td>
<td>446</td>
<td>-27,08</td>
<td>-6,28</td>
</tr>
<tr>
<td>3 year</td>
<td>752</td>
<td>539</td>
<td>699</td>
<td>-39,52</td>
<td>-7,58</td>
</tr>
</tbody>
</table>
FIGURE 4: Mean recorded number of injury crashes per year in the before period and the after period, both for no control for any confounding variable, and with control for trend effects.
FIGURE 5: Mean recorded number of severe crashes per year in the before period and the after period, both for no control for any confounding variable, and with control for trend effects

Discussion

From this study it can be clearly stated that RTM exist and can have a high magnitude. Especially for the injury crashes, high counts of RTM were found. These results are in line with previous studies that found less effective outcomes for the implementation of a traffic safety measure, when RTM was taken into account (for example Erke, 2008; Harwood et al., 2007). In contrary to all injury crashes, the present study found a more limited RTM for the crashes with severe injuries and fatalities. This is in contrast to an earlier study (Gains et al., 2005), which showed that RTM only had a modest effect on all injury crashes, but had an appreciable effect on fatal and serious injury crashes. The difference between the number of severe crashes from the before to after period can, according to the analyses, for a high part be ascribed to the trend effect. It is indeed the case that, both the comparison group, which encompasses all crashes at intersections with regional roads, as all crashes in Flanders (BIVV, 2008) show a higher decrease for the severe crashes compared to all injury crashes during the years 2000-2002. As all injury crashes encompasses both crashes with slightly injured and crashes with dead and severely injured, this difference will have to be due to only a small decrease or even a stagnation of the slightly injured. This is clearly displayed in figure 6, which shows all crashes for the comparison group for which a distinction is made between crashes with only slightly injured, with severely injured, and with dead. From this it is obvious that the crashes with dead, but especially severely injured showed a higher decrease, compared to the crashes with slightly injured. Another possible explanation is that the research group is selected based on the number of all crashes, and no distinction was made according to the severity of the crashes. This research group is a subset of the dangerous spots, that according to the weighting method (a weigh of 1 for slightly injured, 3 for severely injured and 5 for deadly injured) consist of sites that during 1997-1999 had a high number of crashes with slightly injured and a lower number of crashes with severely or fatal injured. As in this study only the sites with the highest number of crashes were selected, it can be expected that especially the sites with crashes with slightly injured will have been selected. The sites that had more severe crashes will be much lower. Subsequently the research group possibly had no extreme number of severe crashes during the before period, and therefore no RTM is found.

The magnitude of the regression from a high number of crashes, to a more average is clearly shown. However, it would be interesting to apply this research also on sites that had a low number of crashes, to see to what extent they return to the higher average number.
FIGURE 6 Percent difference in number of crashes in the comparison group, subdivided among severity of the crashes, compared to the crashes from 2000

Some skeptics stated that taking into account enough before and after data, will lead to little or no RTM (Persaud & Lyon, 2007). A study of Sharma and Datta (2007) found a RTM effect that became insignificant when three or more years of crash data are used in the evaluation of high crash locations. This was not found here, even with the inclusion of three years of B&A data, the RTM was highly present. Nicholson (1988 in Elvik, 1997b) showed that prolonging the before period, will not diminish the RTM, but will reduce it. Also Hauer & Persaud (1983) suggested that a long before period will not eliminate RTM, especially when average annual crash count is relatively small. In contrary, our study found no high differences between the number of years of before and after data that were included. In this study a maximum of three years is taken into account. But when more years would be included, probably the RTM effect would diminish. However on this way the phenomenon RTM is not properly handled. RTM fluctuates over the years, so when too much years are taken into account, the more average number of crashes will be calculated. In most cases the decision to execute a measure is taken on basis of a limited number of years. Taking a couple of years of crash data into account could help to lower the RTM, but on the other hand one have to take care that long before and after periods enlarge the influence of general trends in crashes on the results of a study (Elvik, 1997b).

In a first attempt, the comparison group was selected through a yoked method. For every site in the research group, a site in the neighborhood of this site was selected. However this comparison group showed an increase in the number of crashes which is opposite to the general trend in crashes. This increase could possibly be explained by the fact that the 1014 most dangerous sites, as defined in 2000 by the Flemish government, are not taken up into the comparison group. All the sites at intersections with a regional roads that had a high number of crashes were excluded, and subsequently only the ones with a low number of crashes remained. Probably this led to the selection of a comparison group that showed the opposite direction of RTM, that is a low number of crashes that returns to a more average one. As it is important that the comparison group reflects the general trend, and the general trend shows a decrease in number of crashes, another comparison group was chosen. All sites from the research group are situated at intersections at regional roads, and therefore all crashes at intersections at regional roads were selected. According to Hauer (1997) a good comparison group encompasses next characteristics: a past crash history which closely matches that at the treated sites, a large enough crash count for the effects of random fluctuations to be small, a group of sites unaffected by the safety measure under investigation. The first element was difficult to apply, as the difference in number of crashes
between the research and comparison group was the subject of research. The second element were fairly good applied. And the third element was not applied here. When a comparison group is used instead of a randomly selected control group, there will be inevitably systematic differences between research and the comparison groups (Hirst et al., 2004). The comparison group need to comprise the crashes occurring in the same time period, over the wider area in which the research group is situated, from which the similarity of past crash history can be ensured (Hirst et al., 2004). Hirst et al. (2004) stated that regional or national crashes are an obvious choice. As here all sites from the research group, which are all located at intersections at regional roads, are scattered over the whole area of Flanders, it is assumable that the crashes at all intersections with regional roads in Flanders are included in the comparison group.

Different variables are defined as confounding variables in before-and-after studies (Elvik, 1997b; Hauer, 1997). Although no measure was applied in this study, it was important to control for all those factors, to make sure no false conclusions were made about the magnitude of the RTM. A first important confounding variable is RTM. A second confounding variable is the general trends in the number of crashes. This was properly controlled, through the use of a comparison group. A third confounding variable are changes in traffic volumes. These volumes were not explicitly taken into account, as those were not available for all the research sites. According to Elvik (2002) there is no need to control explicitly for traffic growth. He stated that it is sufficient to include a large comparison group, which he defines as one in which the annual count of crashes is at least several hundreds. He stated that this includes the effect of all factors that may produce changes over time in the long-term expected number of crashes. There is then no need for estimation of the effects of changes in traffic volume statistically, and doing so carries a great risk of double counting. As the comparison group in this article encompasses national wide crashes, which counts for thousands of injury crashes and hundreds of severe crashes, it can be concluded that the comparison group controls for changes in traffic volumes. A fourth confounding variable is the crash migration, due to the implementation of a traffic safety measure. As no adaptations were executed during the whole study period, and everything remained the same, it can be expected that crash migration will not have been occurred. A fifth confounding variable is chance. This is taken into account by the application of a large comparison group. Another confounding variable is the introduction of any other event introduced at the same time as the road safety measure. This was not the case in this study as all the sites remained the same during the after period, and traffic safety measures that were introduced, were more widely implemented, as for at least Flanders.

Summarized we can conclude it is of high importance that RTM is taken into account in future studies that examine the effectiveness of a measure. If not done so, the effectiveness could be overestimated at sites that knew a high crash count during the before period. Controlling for RTM can be done by using the Empirical Bayes method. This method controls both for trend as for RTM. However, different issues are defined as having an effect on the validity of the empirical Bayes results (Persaud & Lyon, 2007). Furthermore this method had some criticism as some skeptics stated that the sophistication and data that are needed to execute this method, is not worth the effort. They stated that alternative, less complex methods can produce equally valid results (Persaud & Lyon, 2007). However besides these weaknesses and the criticisms, the empirical Bayes method is stated as a method that does produce results that are substantially different and more valid than those produced by more traditional methods and is indicated as the best method to control for RTM (Elvik, 2008; Hauer, 1997; Persaud & Lyon, 2007). Also different empirical Bayes estimates are applicable. An overview of those estimates, together with the predictive performance of each of them is given by Elvik (2008). Here five estimates are described and examined, from which was concluded that empirical Bayes-estimates based on crash prediction models showed the lowest prediction errors, and these were defined as the state-
of-the-art approach for observational B&A studies (Elvik, 2008). Next to the empirical Bayes method, also the full Bayes method can be used. For this, different advantages are defined, such as the possibility to use the method for quite complex model forms that are not easily handled in conventional generalized modeling approaches. Also it allows the estimation of valid models with smaller sample sizes. The disadvantage however is that the methodology is quite complex and may require a high level of statistical training (Persaud & Lyon, 2007).

Next to the empirical Bayes method, some other techniques can be defined. An ideal way of avoiding the impact of RTM is selecting sites on a random manner from amongst a specified set of potential sites, with those sites not selected then acting as controls. However this is rarely possible, because of ethical reasons. Maher & Mountain (2009) define an alternative possibility, that is after the sites for treatment are selected on the basis of their crash record in the before period, the implementation of the treatment has to be delayed for some time. This is defined as the lag period. The lag period can be used as an unbiased estimate of the true crash rate before the treatment is applied, and instead of comparing the crashes after with before, the crashes could be compared from after with the lag period (Maher & Mountain, 2009). However this will not be applicable in every situation as often the government or traffic safety organizations, try to implement a measure as soon as possible and a delay can encounter opposition.

References


Elvik, R (2002). The importance of confounding in observational before-and-after studies of road safety measures. Accident Analysis & Prevention, 34, 631-635.


