THE MAGNITUDE OF THE REGRESSION TO THE MEAN EFFECT IN TRAFFIC CRASHES

Ellen De Pauw¹, Stijn Daniels¹, Tom Brijs¹, Elke Hermans¹, Geert Wets¹*

¹Transportation Research Institute
Hasselt University
Wetenschapspark 5
BE-3590 Diepenbeek
Belgium
Tel: +32112691{57, 56, 55, 41, 58}
Fax:+32(0)11 26 91 99

Email: {ellen.depauw, stijn.daniels, elke.hermans, tom.brijs, geert.wets}@uhasselt.be

* Corresponding author

First submitted: August 1, 2012
Revised paper submitted: November 15, 2012

Number of words: 5635
Number of tables: 3
Number of figures:
→ Total number of words: 6385
ABSTRACT

Regression to the mean has been recognized as a phenomenon that influences road safety evaluations and should be accounted for. However, some doubts have risen about the necessity to implement rather sophisticated techniques such as the empirical Bayes method to correct for regression to the mean whereas the use of a sufficient long before-period could reach the same objective. Present study examines the existence and the magnitude of the regression to the mean effect in crash data from 169 intersections for the injury crashes and 143 for the more severe crashes in Flanders-Belgium for whom regression to the mean was likely to occur as they were selected based on their crash history. The presence of a RTM-effect was investigated by comparing the crash numbers of this period with the crash numbers in the next three years, during which no traffic safety measure was applied. Two comparison groups were used. The results demonstrate the existence of a substantial regression to the mean effect in the investigated sample of intersections. The magnitude of the regression to the mean effect is estimated to be almost 9% for injury crashes and 37% for the most severe crashes. From this can be concluded the correction for regression to the mean in evaluation studies is highly recommended in cases when locations are selected based on their crash history. This can be applied through the Empirical Bayes method.

Key words: regression to the mean, empirical Bayes, before-after evaluation
INTRODUCTION
To implement an effective traffic safety policy, there is an undisputed need to evaluate the effectiveness of traffic safety measures. Well-controlled before-and-after (B&A) studies are widely considered to be the most appropriate method to execute such evaluation (1, 2). 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 crashings or injuries whose effects, if not estimated, can be mixed up with the measure being evaluated (1). Variables that are regarded as potentially confounding in observational B&A studies of road safety measures are regression to the mean, long term trends 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 (1, 3, 4). Regression to the mean (RTM) is defined as one of the most important confounding variables (3). Elvik and Vaa (6) defined it as followed: “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. RTM occurs as a result of random fluctuation in the recorded number of accidents around the long-term expected number of accidents”. As often the decision to implement a measure is based on a high crash rate during a relatively short period (e.g. 1 year), it is plausible that the number of crashes will decrease afterwards, irrespective of the measure. In those cases RTM will lead to an overestimation of the treatment effectiveness, when not appropriately taken into account.

BACKGROUND
Several authors analyzed the RTM phenomenon when examining the effectiveness of traffic safety measures, others compared the results of studies that controlled for RTM with studies that did not controlled. Based on several real-world data sets, Hauer and Persaud (7) found at locations selected because of their poor accident history, simple B&A comparisons tend to overestimate the safety effect of a measure, due to the RTM phenomenon. Elvik (1) compared studies that did not controlled for RTM, with studies that did controlled. This analysis showed the magnitude of RTM was estimated to be 20-30% in three out of seven cases, but was of a negligible magnitude in the other four. Harwood et al. (8) applied three B&A methods to evaluate the safety-effects of the installation of added left-turn lanes, right-turn lanes and the extension of the length of existing left- or right-turn lanes. The Empirical Bayes (EB) method, that controls 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 ascribed this difference to the RTM phenomenon. In a study that analyzed the effectiveness of speed and red light cameras (9), the authors also examined the RTM effects on a subset of 216 cameras, through application of the EB method with an external crash model. This study 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 decrease in injury crashes, and three fifths of the observed decrease in more severe crashes. A meta-analysis that examined the effectiveness of red light cameras (10) showed that studies that controlled for RTM and spillover effects, yielded less favorable results, compared to studies that did not controlled these confounding variables. The study concluded that a failure to control for RTM can lead to an overestimation of the effect of RLC (10). Sharma & Datta (11) studied the RTM effect using four B&A evaluation methods (B&A, B&A with comparison group, 2 EB methods) through which the effectiveness of low-cost safety improvements were evaluated. From this study it was found that analyses that did not controlled for RTM, produced similar results to those that did controlled, when three to five years of crash rates were included. The authors concluded that the RTM effect becomes insignificant when three or more years of crash data are applied in the evaluation of high crash locations.

The occurrence of RTM in B&A studies is widely accepted. However, only few studies examined the magnitude of this RTM effect. Furthermore doubt exist concerning the necessity to use sophisticated methods, from which the EB method is stated as one of the most appropriate, whereas others stated the use of many years of pre-treatment data is sufficient to control for RTM (11).

STUDY DESGIN
In order to analyze the RTM effect, a dataset of locations that knew a higher than average number of crash counts was selected. Using a B&A design, these crash rates were compared with the crash rates in the years after, during which no traffic safety measure was implemented. Three B&A methods were applied, which differ in the extent they control for trend and RTM. The magnitude of the RTM effect was examined through a comparison of the results using these three B&A methods.
DATA

To measure the RTM effect, a subset of the data set from the black spot program in Flanders, Belgium was used. To increase traffic safety, the government decided in 2002 to select the most dangerous traffic spots in Flanders and to redesign these locations. A location was selected as ‘dangerous’ when during the years 1997-1999 at least three injury crashes occurred, and a total severity score of 15 was reached. For this score the number of injured were taken into account: every slightly injured (defined as every person that got injured in a traffic crash, but cannot be defined as severely or dead injured) got a weight of 1, every severely injured (every person that needed more than 24 hours of hospitalization as a consequence of a crash) got 3 and every deadly injured (every person that died as a consequence of a traffic crash on the spot or within 30 days after the crash) got a score of 5. A total priority score of minimum 15 was necessary to be selected as a dangerous spot. A total of 1014 black spots were selected, of which the 800 most dangerous were planned to be adapted. Assuming the existence of a RTM phenomenon, it can be expected that locations with this high number of crashes will return to a more average number in the years after, even without the implementation of a traffic safety measure. However, generally the RTM-effect cannot effectively be isolated from other effects since several phenomena (e.g. trend effects, physical measures, changes in traffic volume, random fluctuations, RTM) occur more or less simultaneously. The used dataset was particularly suitable to examine the RTM effect as the locations were based on their crash history in the period 1997-1999 and the adaptations of the locations only started in 2003. Between the period that was used to select the black spots and the starting point of the first adaptations, also called as the lag period (13), no traffic safety measures were implemented on these locations. This period (2000-2002) could subsequently be used, to examine whether RTM occurred.

To analyze the RTM phenomenon, a subset of the black spots with the highest frequency of crashes was selected. A first selection was made for the injury crashes, for which the 200 spots with the highest number of injury crashes in a radius of 100 meter during 1997-1999 were selected. Eventually 169 sites were included in the final research group. Thirty-one spots could not be included, as there was not enough information about the traffic volume data, which was necessary to apply the empirical Bayes method (see further). In order to analyze the RTM-effect in severe crashes, spots with 3 or more crashes with serious injuries or fatalities that occurred during 1997-1999 were selected, which led to a selection of 180 spots. The final research group for severe crashes consisted of 143 locations, as for 37 spots not all necessary data was present.

As stated above, different variables are defined as confounding in B&A studies, from which RTM is one (3, 14). Although no measures were implemented during the research period, it was important to control for all those factors, to make sure no false conclusions were made about the magnitude of the RTM. One of these confounding variables is the general trend in the number of crashes. To account for this variable, a comparison group can be used (1). 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 (=comparison group 1). As a check for robustness, also a second comparison group is used, which encompasses all crashes in Flanders (=comparison group 2). A second confounding variable are changes in traffic volumes. According to Elvik (1) there is no need to control explicitly for traffic growth. He stated it is sufficient to include a large comparison group, with several hundreds of crashes. Such a comparison group includes the effects of all factors that may produce changes over time in the long-term expected number of crashes. In such cases there is no need to estimate the effects of changes in traffic volume statistically, and doing so carries a great risk of double counting (1). As both comparison groups in this article encompass national wide crashes, which counts for thousands of injury crashes and hundreds of severe crashes, it can be concluded that the comparison groups control for general changes in traffic volumes. A third confounding variable is chance, which is taken into account through the application of a large comparison group, several years of B&A crash data and the application of point estimates and confidence intervals. Fourthly, also the introduction of any other event introduced during the study period can be defined as a confounding factors. This was not the case in this study as all the locations remained the same during the total research period, and traffic safety measures that were introduced, were more widely implemented for the total of Flanders. These were subsequently taken into account through the comparison groups.

To examine whether the comparison group is comparable to the research group, the odds ratio (OR) for the crash rates during the before period (1997-1999) can be estimated (15).

\[ \frac{R_t}{C_t} \]  

\[ \frac{R_{t-1}}{C_{t-1}} \]

\[ OR = \frac{R_{t}/R_{t-1}}{C_{t}/C_{t-1}} \]  

\[ R_t = \text{number of crashes in the research group in year } t \]

\[ R_{t-1} = \text{number of crashes in the research group in year } t-1 \]

\[ C_t = \text{number of crashes in the comparison group in year } t \]

\[ C_{t-1} = \text{number of crashes in the comparison group in year } t-1 \]
An OR near to one indicates the comparison group is comparable to the research locations. Maximum standard deviation should not be higher than 0.20. Those analyses are applied on all injury crashes as on crashes with only severely or deadly injured. The results in table 1 show for comparison group 1 and 2, both for all injury crashes as for the severe crashes, the odds ratios are near to 1, and the standard deviations do not exceed 0.20. We can conclude these are appropriate comparison groups.

<table>
<thead>
<tr>
<th>Comparison group 1</th>
<th>Comparison group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury crashes OR (s)</td>
<td>Severe crashes OR (s)</td>
</tr>
<tr>
<td>97-98</td>
<td>0.95 (0.05)</td>
</tr>
<tr>
<td>98-99</td>
<td>1.09 (0.05)</td>
</tr>
<tr>
<td>Average</td>
<td>1.02 (0.05)</td>
</tr>
</tbody>
</table>

**TABLE 1: Odds Ratio (OR) and Standard Deviation per Year, both for Injury and More Severe Crashes**

**METHOD**

To calculate the magnitude of the RTM effect, the crash rates from the after period are compared with the crash rates from the before period, using three different methods. This comparison can be expressed by an index, similar to the index of effectiveness (3), however slightly different here as no measure was applied, and thus it cannot be referred to as ‘effectiveness’. Therefore ‘index of difference’ is a better expression. The proportion of the number of crashes before to after can be expressed by a relative count that indicates how much of the crashes remain during the after period:

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

When the index is equal to one, this indicates there was no difference in the number of crashes. An index lower than one shows that the number of crashes decreased. An index higher than one designate an increase in the number of crashes after, compared to before. In order to examine the magnitude of the RTM effect, three different methods are used to calculate the index of difference, which differ in the extent they control for confounding variables.

At first the naïve B&A method was applied. This method shows how the number of crashes has evolved over the years, without control for any confounding variable. However, even when no measure is applied, it can be expected that the crash rates will change according to the crash trend. To examine which part of the change is due to general crash trend, a second method is applied, which compares the crash rates after with before, taking trend effects into account through a comparison group. As no traffic safety measure were applied at the research locations, 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 still a decrease in the crash rates will be found, even when trend is accounted, this difference can be ascribed to RTM. To confirm these results, a third method was applied, which is the EB method. This method calculates the evolution of crashes and controls, next to trend and chance effects, also for RTM effects (1, 3). When the application of this method leads to a non-significant result, this is a prove of the existence of RTM at the selected sites. Comparing the results of the application of those three methods on the dataset could reveal what was the difference between the results of the analyses for which no confounding variables were controlled, only trend was taken into account and both trend and RTM was controlled. Those three methods were applied on injury and severe crashes, for which three years of before (‘97-’99) and after crash data (‘00-’02) were taken into account.

**Control for RTM**

To control for RTM, Hauer et al. (16) give next equation:

\[
L_{RTM, \text{before}} = w \ast (\mu_{C, \text{before}} \ast km_{L} \ast T_{\text{before}}) + (1-w)\sum_{t=1}^{T_{\text{before}}} L_{t}
\]

With:

\[
L_{RTM, \text{before}} = \text{estimated number of crashes during the before-period in location L after correction for regression for the mean}
\]
\( w \) = the importance of the comparison group
\( \mu_{C, \text{before}} \) = average number of crashes per year in the comparison group during the before period
\( \text{km}_L \) = length of location L in km (intersections get a length of 1km)
\( T_{\text{before}} \) = length of period before the measure
\( L_t \) = number of crashes at location L in year t
\( \text{I} - w \) = the importance of location L.

To calculate the average number of crashes in the comparison group (\( \mu_C \)), two possibilities can be chosen: a comparison group that consist of crash data from comparable locations or a model that estimates the crash rates, based on the characteristics of comparable locations. In this study a model is used, which is also based on the Flemish black spot dataset (17). The dependent value of the model was the number of crashes, based on crash rates of 2000-2003. This period is located after the selection of the black spots, for which crash rates from 1997-1999 were taken into account, and through which the effect of RTM is excluded. Furthermore this period is located before measures were implemented, as the first locations were adapted in 2004. The model was based on locations with varying types of intersection (e.g. both signalized as priority controlled, 3leg and 4leg, urban and rural setting). In this research a model is used which estimates the number of injury and severe crashes through traffic volumes at major and minor roads of the intersection:

\[
E_{\text{injury}}(\lambda) = e^{-1.7131} Q_{\text{Maj}}^{0.3231} Q_{\text{Min}}^{0.2463} \\
E_{\text{severe}}(\lambda) = e^{-3.2138} Q_{\text{Maj}}^{0.3227} Q_{\text{Min}}^{0.2009}
\]

Where
\( E(\lambda) \) = expected annual number of crashes (dependent variable), with \( E_{\text{injury}} \) are all crashes with injured, and \( E_{\text{severe}} \) are all crashes with severe injuries and fatalities
\( Q_{\text{maj}} \) = traffic volume at major road
\( Q_{\text{min}} \) = traffic volume at minor road
\( e \) = natural logarithm = 2.718

A more detailed description of the model can be found in De Ceunynck et al. (17).

The weight (\( w \)) can be calculated through next equation:

\[
w = \frac{1}{1 + \frac{(\mu_{C, \text{before}} \cdot T)}{\text{km}_L}}
\]

With \( k \) is an over dispersion parameter per unit of length (18, 19), which is also calculated through the model.

Control for Crash Trend

Applying control for trend through a comparison group, can be written as an odds ratio:

\[
\frac{L_{\text{after}} / \text{km}_L}{L_{\text{before}} / \text{km}_L}
\]

With:
\( L_{\text{after}} \) = number of crashes on location L during the after period
\( L_{\text{before}} \) = number of crashes on location L during the before period

or estimated number of crashes during the before period after control for RTM (for EB method)

With a 95% confidence interval (CI):

\[
\text{EFF, below limit} = \exp[\ln(\text{EFF}) - 1.96 \cdot s]
\]

\[
\text{EFF, above limit} = \exp[\ln(\text{EFF}) + 1.96 \cdot s]
\]

And a standard deviation (\( s \)) as the root of the variance (\( s^2 \)):
\[ s_i^2 = \frac{1}{\hat{l}_{after}} + \frac{1}{\hat{l}_{before}} + \frac{1}{\hat{c}_{after}} + \frac{1}{\hat{c}_{before}} \]  

[9]

Meta-Analysis

Next to the individual evaluation per location, the total effect across different locations can be calculated by means of a fixed effects meta-analysis, which results in one overall effect estimate and in more statistically reliable results \((15)\). Every location within the meta-analysis gets a weight, which is the inverted value of the variance.

Subsequently locations at which many crashes occurred, are given a higher weight.

\[ w_i = \frac{1}{\sigma_i^2} \]  

[10]

Supposing that the measure is executed at n different places, the weighted mean index of difference of the measure over all places is:

Overall index of difference = \(\exp\left[\frac{\sum_{i=1}^{n} W_i \ln(\text{EFFI})}{\sum_{i=1}^{n} W_i}\right] \)  

[11]

The estimation of a 95% confidence interval is:

\[ 95\% \text{ CI} = \exp\left[\frac{\sum_{i=1}^{n} W_i \ln(\text{EFFI})}{\sum_{i=1}^{n} W_i}\right] \pm 1.96 \times \frac{1}{\sqrt{\sum_{i=1}^{n} W_i}} \]  

[12]

RESULTS

At first all three methods (naïve B&A, B&A with comparison group, EB method) were applied on every black spot, separately for injury crashes and more severe crashes, taking three years of B&A data into account. Afterwards an overall estimation for all locations (169 for injury crashes, 143 for severe crashes) was executed with each of the three methods. The results are shown in table 2. For the injury crashes a significant decrease from before to after of 13% in the number of crashes was found, using the naïve method. This is a quite high decrease as no specific measures were applied at these locations. As a consequence those decreases are attributable to confounding variables, such as trend effects and RTM. When trend and chance are taken into account, using the B&A method with comparison group 1 (all crashes at intersections at regional roads), the results show an index that is still lower than 1, but higher compared to the results of the naïve method. This indicates that the decrease from before to after is lower when trend is controlled, and subsequently that trend accounts for a part of the recorded decrease in injury crashes. However, even with this correction a nearly significant decrease of 5% was found. Using comparison group 2, that encompasses all crashes in Flanders, a decrease of 9%, significant at the 5% level was found. As the most important confounding variables are controlled and no measure was implemented, this difference can subsequently be attributed to the RTM effect. To confirm this statement, the EB method is applied. When indeed RTM occurred it can be expected that analyses using the EB method -since it corrects for RTM- will show non-significant changes in the number of crashes. As from table 2 can be seen, the application of the EB method resulted in a non-significant index of 1.03 with comparison group 1, and 0.99 with comparison group 2. The results using the EB method indicate that there is no significant difference between the recorded number of crashes after and the estimated number of crashes before, when trend and RTM are controlled. This confirms the presence of RTM and the suitability of the EB method to correct for its existence.

For the severe crashes, the naïve B&A method shows a high decrease in crash rates of 36%. When the trend is taken into account, this resulted in a higher index of difference. However, still a significant decrease of 21% with comparison group 1 and 24% with comparison group 2 is found. From this can be concluded that RTM is also present for severe crashes. The application of the EB method, which found a non-significant increase of 6 to 11%, confirms this statement.
TABLE 2: Results (Index of difference, [95% Confidence Interval])

<table>
<thead>
<tr>
<th></th>
<th>Naive B&amp;A</th>
<th>B&amp;A with comparison group</th>
<th>EB method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index [95% CI]</td>
<td>Index [95% CI]</td>
<td>Index [95% CI]</td>
</tr>
<tr>
<td></td>
<td>Comparison group 1</td>
<td>Comparison group 2</td>
<td>Comparison group 1</td>
</tr>
<tr>
<td>Injury crashes</td>
<td>0.87 [0.82; 0.92]</td>
<td>0.95 [0.90; 1.01]</td>
<td>0.91 [0.86; 0.97]</td>
</tr>
<tr>
<td>Severe crashes</td>
<td>0.64 [0.56; 0.74]</td>
<td>0.79 [0.68; 0.91]</td>
<td>0.76 [0.65; 0.88]</td>
</tr>
</tbody>
</table>

DISCUSSION

Table 3 gives a more detailed analysis of the magnitude of the RTM effect. This table shows the proportional change in the crash rates from before to after (see column 3), and the proportional change in crash rates from before to after, attributable to the RTM phenomenon (column 4). For these calculations the results with comparison group 1 are used, as it can be expected that crashes at intersections at regional roads are more similar with the crashes at the research locations, compared to all crashes in Flanders, which encompass for example also crashes at road sections. As the RTM effect is a group phenomenon, it is not reasonable to compute the deviation for each location separately (11). Therefore those results were grouped into various ranges of crash frequencies from 1997-1999 (see column 1), for which an average value for the group in that range was calculated. In the last row the average over all locations is calculated. The results show, both for the injury and the more severe crashes, the general change in the number of crashes is negative, which indicates a decrease in crashes. Also the changes attributable to RTM is negative, for which the magnitude ranges from 1.5 to 17.7%, with an average of 8.7% for all injury crashes. For the severe crashes a decrease of 20.6 to 54%, with an average of 36.9% is found. Based on the total change in crash rates and the change attributable to RTM, column 5 shows which part of the observed decrease is represented by the RTM effect. From this can be seen that, the part of the observed decrease attributable to RTM raises proportionally with the mean number of crashes from the before period (as shown in the first column). For the injury crashes 18 to 66%, with an average of 51% of the observed decrease from before to after can be ascribed to the RTM phenomenon. For the severe crashes this ranges from 59 to 81%, with an average of 74%. From these results can subsequently be concluded the RTM effect for severe crashes is higher compared to all injury crashes.

TABLE 3: Detailed Analysis of the Magnitude of the RTM Effect, Using Comparison Group 1

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Column 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean nr crashes in before period</td>
<td>Nr of locations in that group</td>
<td>Total change in crash rates (%)</td>
<td>Change in crash rates attributable to RTM (%)</td>
<td>Part of the observed decrease attributable to RTM (%)</td>
</tr>
<tr>
<td>All injury crashes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-3.9</td>
<td>26</td>
<td>-8.1</td>
<td>-1.5</td>
<td>18.0</td>
</tr>
<tr>
<td>4-4.9</td>
<td>65</td>
<td>-14.9</td>
<td>-6.1</td>
<td>41.1</td>
</tr>
<tr>
<td>5-5.9</td>
<td>39</td>
<td>-13.8</td>
<td>-5.5</td>
<td>40.2</td>
</tr>
<tr>
<td>6-6.9</td>
<td>15</td>
<td>-23.1</td>
<td>-14.6</td>
<td>63.4</td>
</tr>
<tr>
<td>7-7.9</td>
<td>11</td>
<td>-15.7</td>
<td>-6.8</td>
<td>43.1</td>
</tr>
<tr>
<td>≥8</td>
<td>12</td>
<td>-26.9</td>
<td>-17.7</td>
<td>65.9</td>
</tr>
<tr>
<td>Average</td>
<td>-17.1</td>
<td>-8.7</td>
<td>-1</td>
<td>51.0</td>
</tr>
<tr>
<td>Severe crashes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>74</td>
<td>-34.9</td>
<td>-20.6</td>
<td>59.1</td>
</tr>
<tr>
<td>1.01-1.5</td>
<td>30</td>
<td>-38.8</td>
<td>-28.3</td>
<td>72.9</td>
</tr>
<tr>
<td>1.51-2</td>
<td>24</td>
<td>-59.4</td>
<td>-44.6</td>
<td>75.1</td>
</tr>
<tr>
<td>≥2</td>
<td>15</td>
<td>-66.5</td>
<td>-54.0</td>
<td>81.2</td>
</tr>
<tr>
<td>Average</td>
<td>-49.9</td>
<td>-36.9</td>
<td>-</td>
<td>73.9</td>
</tr>
</tbody>
</table>

DISCUSSION

Present study shows that the number of crashes at locations that were selected based on their high crash counts returns to a more average number during the years after, even when no measures were applied and possible other
influencing factors were eliminated. This decrease can subsequently be attributed to the RTM effect. The magnitude of this effect is estimated to be almost 9% for all injury crashes, which represents more than half of the observed decrease from before to after. For severe crashes a decrease of 37%, attributable to RTM is found, which count for almost three fourths of the observed decrease. Even when three years of B&A data are taken into account, high decreases can be found. This is in contrast with the study of Sharma and Datta (11), who stated that the RTM effect becomes insignificant when three or more years of crash data are applied in the evaluation of high crash locations. Furthermore it can be concluded that the RTM effect is higher for the more severe crashes compared to all injury crashes. This is in line with an earlier study (9), which showed that RTM only had a modest effect on all injury crashes, but had an appreciable effect on fatal and serious injury crashes.

A problem that especially needs attention in current study, and more generally in all B&A studies that examine the effectiveness of traffic safety measures, is the occurrence of zero crashes in the before or after period. In current research, no zero counts appeared for all injury crashes, but for the severe crashes 33 locations had zero counts during the after period. Elvik (21) described these zero counts as a problem for three reasons. He stated it is highly implausible that the true long term mean number of accidents at any location is zero. Secondly zero counts suggest that a safety treatment could be either a hundred percent crash reduction (if there was a positive count before and a zero count after) or an infinite increase in the number of crashes (if there was a zero count before and a positive count after), both of which are highly implausible. Thirdly, zero counts have to be adjusted when in a meta-analysis a statistical weight is to be assigned to each result. A conventional method to solve this problem, is to add 0.5 to each of the four variables, as shown in equation [7]. However previous research indicated that applying such a factor can lead to deviant results in meta-analyses (20). Subsequently it is an important initiative to explore this problem more thoroughly in future.

Summarized we can conclude it is of high importance that RTM is taken into account in future B&A studies that examine the effectiveness of a traffic safety measure. If not done so, the effectiveness is likely to be overestimated at locations with a higher than average crash count during the before period. Controlling for RTM can be done through application of the EB method. This method controls for different confounding variables, including RTM. However, this method had some criticism, as skeptics stated that the sophistication and data that are needed to execute this method is not worth the effort. They argued that alternative, less complex methods can produce equally valid results (12). Besides these weaknesses and criticisms, the EB method is stated as a method that does produce results that are substantially different and more valid than those generated by more traditional methods, and is indicated as the best to control for RTM (3, 12, 22). Different EB estimates are applicable, for which previous research found the EB-estimates based on crash prediction models showed the lowest prediction errors, and were defined as the state-of-the-art approach for observational B&A studies (22). Next to the EB method, also the full Bayes method can be used. For this method 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 (12). Next to the EB method, some other techniques can be defined. An ideal way of avoiding the impact of RTM is selecting locations on a random manner from amongst a specified set of potential sites, with those locations that are not selected then acting as controls. However this is rarely possible, because of ethical reasons. Maher and Mountain (13) defined an alternative possibility, that is to postpone the implementation of the treatment for some years, after the locations are selected on the basis of their crash record in the before period. 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 (13). 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.

CONCLUSIONS

Based on these results, the following conclusions can be made:

- This study demonstrates the existence of a substantial regression to the mean effect in a sample of intersections that were particularly prone to this effect.
- The magnitude of the regression to the mean effect in the investigated sample is estimated to be almost 9% for crashes with at least a slight injured and 37% for severe crashes.
- The Empirical Bayes method effectively corrects for regression to the mean.
Correction for regression to the mean in evaluation studies is highly recommended in cases when locations are selected based on their crash history.
REFERENCES