Accident Analysis and Prevention 130 (2019) 22–29

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Effects of methodological decisions on rainfall-related crash relative risk estimates

Alan W. Black*, Gabriele Villarini

IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA, USA

ARTICLE INFO

Keywords:
Rainfall
Relative risk
Odds ratio
Crash reports

ABSTRACT

Numerous studies have examined the influence of rainfall on the relative risk of crash, and they all agree that rainfall leads to an increase in relative risk as compared to dry conditions; what they do not agree on is the magnitude of these increases. Here we consider three methodological decisions made in computing the relative risk and examine their impacts: the inclusion or exclusion of zero total events (where no crashes occur during event or control periods), the temporal scale of analysis, and the use of information on pavement and weather conditions contained with the crash reports to determine relative risk. Our analyses are based on several years of data from six U.S. states (Arkansas, Georgia, Illinois, Maryland, Minnesota and Ohio). Zero total events in the context of weather related crash studies typically provide no information on the actual crash odds and greatly alter the distribution of relative risk estimates and should be removed from the analysis. While the use of a daily time step provides an estimate of relative risk that is not significantly different from an hourly time step for the majority of rural counties in our study area, the same is true of only 39% of the urban counties. Finally, the use of pavement and weather condition information from the crash reports results in relative risk estimates that are lower than the standard approach, however this difference decreases as rainfall totals increase. By highlighting the influence of methodological choices, we hope to pave the way towards the potential reduction in uncertainties in weather-related relative risk estimates.

1. Introduction

Inclement weather conditions such as fog, smoke, and dust (Abdel-Aty et al., 2011), frozen precipitation (e.g., Andrey, 2010; Andrey et al., 2013; Black and Mote, 2015a, b; Mills et al., 2011), and rainfall (e.g., Andrey, 2010; Andrey and Yagar, 1993; Andrey et al., 2003, 2013; Black et al., 2017) have been shown to increase the relative risk — the risk of motor vehicle crash, injury, or fatality relative to periods when inclement weather is absent. These hazards lead to elevated relative risk of crash due to a combination of reduced visibility and loss of friction between tires and the road surface. The studies that focus on rainfall-related crashes agree that rain leads to increased relative risk but vary widely in their estimates of the magnitude of the increase. Meta-analysis of studies of rainfall-related crash relative risk in the United Kingdom, Canada, and the United States found that rainfall led to a 31–111% increase in overall automobile crash rate, with injury crash rates increasing by anywhere from 28% to 70% (Qiu and Nixon, 2008). The wide range of crash relative risk estimates are both a product of meteorological and non-meteorological factors that influence risk; however, Jaroszweski and McNamara (2014) suggest that the greatest impact may be due to methodological choices used in the analysis.

There are a number of methodological factors that will influence the final estimate of crash relative risk, including the spatial and temporal scales used in the analysis and the methods used to calculate relative risk. Previous studies of weather related crash relative risk have used a variety of spatial and temporal scales (Table 1), which are typically dictated by the availability of either the meteorological or motor vehicle crash data required for this type of analysis. Most studies (e.g., Andrey, 2010; Andrey and Yagar, 1993; Andrey et al., 2003, 2013; Hambly et al., 2013; Mills et al., 2011) use point measurements of rainfall from a meteorological station to represent the weather conditions for a city or county that may be hundreds or thousands of square kilometers in size (Jaroszweski and McNamara, 2014). As a result, the precipitation values measured may not truly represent the meteorological conditions experienced at the location of a motor vehicle crash. The use of radar data or other gridded precipitation data with high spatial resolution can mitigate this issue, improving confidence that precipitation occurred at a crash location and in the amount of rainfall accumulated at that site (Jaroszweski and McNamara, 2014). Further,

* Corresponding author at. Present address: Southern Climate Impacts Planning Program and Department of Geography & Anthropology, Louisiana State University, Baton Rouge, LA, USA.

E-mail address: ablack@lsu.edu (A.W. Black).

https://doi.org/10.1016/j.aap.2018.01.023
Received 14 June 2016; Received in revised form 5 September 2017; Accepted 17 January 2018
Available online 23 April 2018
0001-4575/

Table 1
Previous studies of rainfall related crash relative risk, the source of meteorological data and temporal scale used in each, and the estimates of relative risk found.

Author/Year	Source of Meteorological Data	Temporal Scale of Analysis	Relative Risk Estimates
Andrey (2010)	Point Data from Meteorological Station(s)	Six-hourly	1.72
Andrey and Yagar (1993)	Point Data from Meteorological Station(s)	Hourly	1.70
Andrey et al. (2013)	Point Data from Meteorological Station(s)	Six-hourly	1.60
Andrey et al. (2003)	Point Data from Meteorological Station(s)	Six-hourly	1.43–1.67
Black et al. (2017)	Gridded Meteorological Data	Daily	1.10
Hambly et al. (2013)	Point Data from Meteorological Station(s)	Daily	1.22
Mills et al. (2011)	Point Data from Meteorological Station(s)	Six-hourly	1.76

the use of gridded data can allow crash relative risk assessment in rural areas or in any location where weather station data are unavailable.

The temporal scale of the analysis also has an effect on relative risk estimates. For example, the relative risk of crash during inclement weather in Vancouver were 20% greater when using daily data and 90% higher when using six-hourly data as compared to dry periods (Hambly et al., 2013). Previous studies of rainfall-related crashes have used a variety of time steps including monthly (e.g., Eisenberg, 2004), daily (e.g., Eisenberg, 2004; Hambly et al., 2013; Keay and Simmonds, 2005), and hourly (e.g., Andrey and Yagar, 1993). Rainfall can occur over very short durations, so it is possible that any time step could capture both rainfall and dry periods leading to an underestimation of relative risk (Hambly et al., 2013; Brodsky and Hakkert, 1988). Use of a shorter time step can mitigate this issue, but it is presumed that short-term adaptations in driver's behavior also influence the relative risk of crash, and that these adaptations can exert a greater influence on the relative risk estimates when a shorter temporal unit is used (Jaroszweski and McNamara, 2014; Elvik, 2006). Despite these possible concerns, it is accepted that the use of hourly precipitation data results in the most representative measure of rainfall for the estimation of crash relative risk (Jaroszweski and McNamara, 2014). However, no study has systematically assessed the difference in crash relative risk that results from changing the temporal scale of analysis: would we obtain statistically different results if we used two-, three-, four-hourly, or longer rainfall accumulations? Put in a different way: what is the largest temporal rainfall aggregation that provides results that are not statistically different from what we obtain at the hourly scale? Moreover, would we obtain the same answer for urban and rural areas? These are some of the research questions we will address in this work.

In many cases, crash reports contain information about the weather and pavement conditions at the time of the crash. Some studies (Andrey and Yagar, 1993; Andrey et al., 2003) have suggested using this information as a way to identify collisions that occur during rainfall and dry periods in the absence of meteorological data. Other research suggests that while crash, injury, and fatality counts within the crash reports are generally reliable, weather and roadway information are not (Brodsky and Hakkert, 1988; Jaroszweski and McNamara, 2014; Mills et al., 2011). To explore the possibility of calculating crash relative risk directly from the information in the reports, relative risk is calculated in three ways for each period with precipitation as identified from the high spatial and temporal resolution rainfall data. The first method assumes that rainfall occurs for the entire duration of a temporal period (i.e. one hour) where rainfall is reported by meteorological observations, and that all crashes during that time period occurred in rain (subsequently referred to as the standard approach). While most studies of rainfall crash relative risk use this method (e.g., Andrey, 2010; Andrey et al., 2003, 2013; Black and Mote, 2015b; Brodsky and Hakkert, 1988; Jaroszweski and McNamara, 2014; Mills et al., 2011), it is possible that both rainfall and dry periods could occur during a single hour and it is therefore possible that not every crash during the period occurred during rain. In the second approach, relative risk is again calculated for each period with precipitation as identified from the meteorological data. However, rather than assuming all crashes during the period were related to rainfall, this method only considers crashes where the weather conditions field on the crash report indicates rainfall was occurring (subsequently referred to as the rain in reports approach). The final method is similar to the second, but only considers crashes where wet pavement conditions were reported on the crash report (subsequently referred to as the wet pavement approach). Each approach was considered independently, although it is common for rainfall and wet pavement conditions to be reported together. It is assumed that the rain in reports and wet pavement approaches will result in lower relative risk estimates than the standard approach, but it is unknown how much lower the estimates will be and if the differences will be statistically significant. Therefore, this work will assess that relationship.

While we have so far focused on issues related to the data, either by considering the spatial and temporal aggregation or the attributes in the crash reports, there are other methodological issues that could have large impacts in the calculation of the relative risk, but that have received little attention in the literature so far. Two approaches are used by most studies to estimate weather related crash relative risk. The first is an approach involving statistical modeling (e.g., Bergel-Hayat et al., 2013; Eisenberg, 2004), which encompasses different methods, including negative binomial regression, general linear models, and vector autoregressive models, among others. Our analysis, on the other hand, will focus on the matched pair approach (Andrey, 2010; Andrey et al., 2003, 2013; Black and Mote, 2015b; Black et al., 2017; Hambly et al., 2013; Jaroszweski and McNamara, 2014; Keay and Simmonds, 2005; Mills et al., 2011; Sherretz and Farhar, 1978). In this approach, specific time periods such as a day, an hour, or a sub-daily aggregation of several hours (e.g., 3 or 6 h) where precipitation occurs in the study region are paired with a matching time period where precipitation was absent. Typically, matches are made such that the control period without rainfall is at the same clock time and either one week before or after the rainfall event to control for time sensitive factors such as traffic volume and light conditions, with the assumption that these remain relatively constant on a week to week basis. After making matches, the number of crashes, injuries, and fatalities are tabulated for event and control periods, which are then used to estimate crash relative risk.

Studies based on the matched pair methodology have used a variety of approaches to estimate crash relative risk. Some studies use a simple approach to calculate relative risk where the total number of crashes, injuries, and/or fatalities during events are divided by the total number during the control periods, resulting in a single value (Andrey, 2010; Andrey et al., 2003). Other studies use the odds ratio approach (Andrey et al., 2013; Black and Mote, 2015b; Black et al., 2017; Elvik et al., 2009; Hambly et al., 2013; Johansson et al., 2009; Mills et al., 2011), where the number of crashes during events and controls are used to calculate the odds of crash during rainfall and during dry periods. The odds for each event/control pair for a specific location are combined to arrive at the relative risk estimate, and the 95% confidence interval is then calculated (additional details about the odds ratio calculation can be found in Section 2.2.1 and are not duplicated here). However, the odds ratio cannot be calculated if there are zero collisions during both the rainfall and control periods. In medicine, epidemiology, public health, and biostatistics where the odds ratio is commonly used, there is

Table 2
States used in this analysis and the corresponding period of record.

State	Period of Record	
	Start	End
Arkansas	1998	2010
Georgia	1996	2008
Illinois	1996	2010
Maryland	2000	2008
Minnesota[a]	1996	2010
Ohio	1996	2010

[a] Minnesota data unavailable for 2003.

substantial debate about the inclusion of instances where the number of cases in the event and control are equal to zero, referred to as "zero total event" cases (e.g., Parzen et al., 2002; Sweeting et al., 2004). For example, Sweeting et al. (2004) argue that zero total event cases should be removed as they provide no information about the odds, while Parzen et al. (2002) contend that the odds in zero total event cases essentially depends on the ratio of the sample sizes between the event and control, and therefore should be included. The inclusion of zero total events will have an impact on the overall estimate of relative risk, but no studies of weather related crash relative risk that use the odds ratio approach have directly assessed this effect. Further, these studies do not contain any discussion of zero total event cases and if they were included or excluded from their relative risk estimates.

In this work, we examine the effect of various methodological changes on the estimates of crash relative risk based on data for six U.S. states. Relative risk is first calculated in two ways — using the odds ratio approach with and without zero total events included — using rainfall data with high spatial (~ 4 km) and temporal (hourly) resolution. Data are then aggregated to periods ranging from 1 to 24 h to examine the difference in relative risk estimates at different temporal resolutions. Finally, collision relative risk is again calculated using solely the information on pavement and weather conditions contained in the crash reports and compared to the results of the standard analysis that assumes all crashes during time period are related to rainfall. In summary, the research questions we wish to addres are:

- What effect does the inclusion or exclusion of zero total events have on the overall estimate of relative risk?
- Is there a significant difference in crash relative risk estimates between the daily and hourly scales of analysis?
- How do relative risk estimates from the standard method compare to those calculated using the rain in reports or wet pavement approaches?

2. Data and methods

2.1. Data sources

To assess crash relative risk and the effect of methodological choices on relative risk estimates, both meteorological data and crash data are required. The source of rainfall data for this study is the National Centers for Environmental Prediction (NCEP) hourly Stage IV multisensor precipitation dataset (Lin and Mitchell, 2005). The Stage IV dataset provides gridded precipitation values in millimeters at a ~ 4 km × 4km resolution across the United States. The high spatial and temporal resolution of this product allows for precipitation estimates that are more representative of actual rainfall over a large area as compared to using data from meteorological stations located at a single point (Jaroszweski and McNamara, 2014). Motor vehicle crash data were obtained from the National Highway Traffic Safety Administration's (NHTSA) State Data System (SDS). SDS data contain information on the number of property damage only (PDO) crashes, injuries, and fatalities due to traffic collisions and are coded directly from police reports in the 32 states which participate in the system. Although there are some limitations associated with this dataset, primarily that each state has a varying period of record and some states are missing data for one or more years, the SDS dataset is the only comprehensive dataset which contains information on all reported automobile crashes, injuries, and fatalities, and serves as the source of crash data for this analysis.

2.2. Methodology

The odds ratio calculation is the underlying method used to estimate relative risk for each of the methodological changes we wish to assess in this study. As such, Section 2.2.1 addresses the overall odds ratio calculation, while Section 2.2.2 discusses the inclusion or exclusion of zero total events, Section 2.2.3 explores the effects of temporal aggregation, and Section 2.2.4 examines crash relative risk estimates obtained by using weather or pavement information from the crash reports.

2.2.1. Odds ratio calculation

Gridded Stage IV rainfall and crash data for six states (Arkansas, Georgia, Illinois, Maryland, Minnesota, Ohio) from SDS were obtained for the years 2002–2010; however, due to the previously mentioned temporal limitations in the SDS, the exact period of record used for each state varies (Table 2). In addition, the analysis was restricted to the period from 1 May to 30 September for each year to reduce the possible influence of winter precipitation on crashes. While the Stage IV precipitation data are available at high spatial resolution, the SDS data are only consistently available at the county level. To address the mismatch in spatial resolution between the datasets, precipitation values were aggregated to the county level based on the area weighted mean value of the Stage IV gridpoints within each county. Dates and times of the crash reports were adjusted temporally so that both the precipitation and traffic data were in Coordinated Universal Time (UTC). Crash and precipitation data were then joined, resulting in a dataset containing precipitation values and the number of crashes, injuries, and fatalities for each hour in the six study states, which was used in the subsequent analysis.

A matched pair design was adopted for the study, where each hour with measurable precipitation in a county ($\geq$ 0.254 mm, 0.01 in.) was paired with a control hour in the same county where precipitation was absent. As in previous studies that used this approach (e.g., Andrey, 2010; Andrey et al., 2003, 2013; Black and Mote, 2015b; Black et al., 2017; Brodsky and Hakkert, 1988; Jaroszweski and McNamara, 2014; Mills et al., 2011), matches were made such that control hours were exactly one week before or one week after the precipitation hour to control for factors such as daily traffic volume, light conditions, and other time sensitive factors, with the assumption that daily and day of week travel patterns are similar over time. If a precipitation event was unable to be matched to a control, it was excluded from further analysis.

Once the number of crashes, injuries, and fatalities were tabulated for events and controls within a county, estimates of the relative risk of crash, injury, and fatality, and their 95% confidence intervals, were calculated using the odds ratio approach. In the odds ratio approach (e.g., Black and Mote, 2015b; Black et al., 2017; Hambly et al., 2013; Johansson et al., 2009; Mills et al., 2011), each event and control pair is used to calculate the odds of a crash, represented as a ratio of the probability that an event will occur to the probability that the event will not happen (Fleiss et al., 2003). The odds ratio (or approximate relative risk) represents the odds of a crash, injury, or fatality in a given county during an event period to the odds of a crash, injury, or fatality during a control period. This can be expressed using the following equation for any of the ith matched pairs ($i = 1 \ldots g$):

$$OR_i = \frac{(A_i/C)}{(B_i/D)} \quad (1)$$

where A_i is the number of collisions, injuries or fatalities during an hour with rainfall, B_i is the number of collisions, injuries, or fatalities during the matched control hour, while C and D are the number of safe outcomes during the rainfall hour and dry hour respectively. The values of A_i and B_i are contained in the traffic data, while C and D must be estimated. Since there are thousands of vehicle trips and driving maneuvers on average in each study county every hour that do not result in a crash, the values of C and D are very large. Recent studies (Black and Mote, 2015b; Black et al., 2017; Mills et al., 2011) suggest a value of 1,000,000 for C and D, concluding that these values can be set in somewhat subjective manner and that the overall relative risk is not sensitive to the values chosen for C and D when both values are equal.

Once the odds ratio is calculated for an event-control pair, it is then log-transformed so that its distribution can be approximated by a normal distribution. The odds ratio is then assigned a weight inversely proportional to its variance based on the fixed-effects model for combining estimates of risk (Black and Mote, 2015b; Black et al., 2017; Hambly et al., 2013; Johansson et al., 2009; Mills et al., 2011). The variance of the logarithm of the odds ratio is:

$$v_i = \frac{1}{A_i} + \frac{1}{B_i} + \frac{1}{C} + \frac{1}{D} \quad (2)$$

An issue arises when A_i, B_i, C, or D is equal to zero because this would result in a division by zero when calculating the variance. In this study, C and D are fixed and not equal to zero but A_i or B_i can be zero if there were no crashes, injuries, or fatalities observed in either the event or control hour. To counter this, a continuity correction factor is added to A_i, B_i, C, or D (Fleiss et al., 2003; Sweeting et al., 2004). There are a number of methods that can be used to determine the continuity correction (e.g., Parzen et al., 2002; Sweeting et al., 2004), but the most common one is to add 0.5, which results in less bias than the addition of any other constant (Gart and Zweifel, 1967; Haldane, 1955).

The statistical weight of each event-control pair is calculated:

$$w_i = \frac{1}{v_i} \quad (3)$$

The weighted mean odds ratio and overall estimate of relative risk of crash based on a set of g matched pairs, where y_i is the logarithm of the odds ratio and 'exp' is the exponential function, is calculated:

$$\bar{y} = \exp\left(\frac{\sum_{i=1}^{g} w_i y_i}{\sum_{i=1}^{g} w_i}\right) \quad (4)$$

Finally, the standard error of the relative risk estimate is used to calculate the 95% confidence interval for the weighted mean odds ratio (Elvik et al., 2009; Fleiss et al., 2003; Johansson et al., 2009):

$$95\%\ \text{Confidence Interval} = \exp\left[\left(\frac{\sum_{i=1}^{g} w_i y_i}{\sum_{i=1}^{g} w_i}\right) \pm \frac{1.96}{\sqrt{\sum_{i=1}^{g} w_i}}\right] \quad (5)$$

Values of relative risk greater than 1 indicate an increase in risk of vehicle crash during rainfall, while values less than 1 represent a decrease in risk. It is important to note that the relative risk estimates produced by the odds ratio method are not intended to assess the absolute risk of collision (Andrey et al., 2003) or the risk to individual drivers (Andrey and Yagar, 1993), and generally produce a conservative estimate of relative risk (Andrey and Yagar, 1993; Andrey et al., 2003; Brodsky and Hakkert, 1998; Eisenberg, 2004).

2.2.2. Inclusion or exclusion of zero total events

To test the possible influence of zero total events on the relative risk estimates, the odds ratio was first calculated with zero total events included, and then re-calculated with the zero total events removed. In summary, the relative risk estimates based on the hourly data were calculated in two ways for comparison:

1 The odds ratio approach described above, including matched pairs where no crashes were observed during the event and during the control (i.e. pairs where A_i and B_i in Eq. 1 both equal zero)
2 Using the same odds ratio approach but excluding the matched pairs where no crashes occurred during the event and during the control.

Once the impact of including or excluding zero total events is known, those events will either be included or excluded from the following analyses of temporal aggregation and relative risk estimates from the crash reports.

2.2.3. Effects of temporal aggregation

While several studies suggest that it is preferable to use an hourly time step when estimating crash risk (Jaroszweski and McNamara, 2014; Hambly et al., 2013), it is not always possible to obtain both meteorological and crash data at a high temporal resolution. For example, while there are over 8000 stations in the United States that report daily rainfall values, just 2500 stations report hourly totals. Further, projections of crashes under changing climate are generally restricted to the daily scale (e.g., Hambly et al., 2013) due to temporal limitations in global climate model output. Given these factors, it is important to assess the effect of temporal aggregation on risk estimates. To determine the effect of the temporal scale of analysis on relative risk estimates in each county, precipitation totals and the number of crashes were aggregated for each n hourly time step between 1 h and 24 h (i.e., n = 1 h, 2 h, …, 24 h). As with the hourly analysis described above, each time aggregation with precipitation was matched to an aggregation of the same length where precipitation was absent, and relative risk estimates calculated using the odds ratio approach with zero total events removed. The overall goal was to determine for each county what was the largest temporal aggregation that would result in odds ratio estimates that were not statistically different (at the 5% level) from those from the hourly resolution. If there is no statistically significant difference in the relative risk estimate calculated using a 24 h (i.e., daily) aggregation and the relative risk determined from the 1-hour time step, then daily meteorological and crash data are all that is needed to adequately estimate changes in crash relative risk due to rainfall for the county in question.

2.2.4. Estimates of relative risk from crash reports

Finally, relative risk was calculated using the rain in reports and wet pavement approaches to determine if these methods produced relative risk estimates that are significantly different from the results of the standard approach. For the rain in reports approach, the number of crashes and injuries were tabulated for each matched event-control period previously identified from the meteorological data where the crash report indicated rainfall during the event. The wet pavement crashes were compiled in a similar manner, but for crashes where the report identified a wet pavement condition during the event. Relative risk estimates were then calculated for both the rain in reports and wet pavement approaches using the odds ratio method described above and with zero total events removed; these estimates were then compared to the standard approach which assumes all crashes during a period with precipitation are related to the precipitation. Relative risk estimates for all three methods were also stratified by precipitation thresholds, including hours with rainfall larger than or equal to 0.254 mm (0.01 in.), 2.54 mm (0.1 in.), 6.35 mm (0.25 in.), and 12.7 mm (0.5 in.).

3. Results and discussion

3.1. Inclusion or exclusion of zero total events

Our analysis found that the vast majority of matched event-control pairs defined by the rainfall data resulted in zero total events. Over

800,000 matched pairs were found with precipitation of at least 0.254 mm during the event and no precipitation during the control. However, just over 25% (202,034) of those matched pairs recorded a crash during either the event or the control period. The percentage of matched pairs with one or more crashes increased modestly at higher precipitation thresholds to 27%, 28%, and 29% for hours receiving at least 2.54 mm, 6.35 mm, and 12.7 mm of rain respectively.

While the overall number of matches decreased by 75% when excluding zero total events, the reduction was not evenly split across urban and rural counties. The number of matches in the 171 urban counties in the six study states based on the "urbanized area" designation from the U.S. Census Bureau (2015); Supplemental Fig. S1) decreased by 50% when excluding zero total events, while the number of matches in rural interstate counties (those not designated as an "urbanized area" that also contain any length of interstate highway; 114 counties) decreased 83% and those in rural counties (250 counties that do not contain "urbanized areas" or interstates) dropped by 89%. On average, each urban county had the number of matched pairs reduced by 763 when removing zero total events, while rural interstate counties decreased by 1261 and rural counties saw a reduction of 1309.

The presence or absence of zero total events has a significant impact on the overall relative risk calculation. Crash relative risk was estimated to be 1.184 (95% confidence interval: 1.179–1.188) when including zero total events and 1.386 (95% confidence interval: 1.379–1.392) after their exclusion, a difference of around 20%. The difference between relative risk estimates with zero total events included and excluded increases with increasing hourly rainfall totals. Relative risk with zero total events excluded is 27%, 31%, and 35% higher than with zero total events included for hours receiving at least 2.54 mm, 6.35 mm, and 12.7 mm of rain respectively.

The distribution of relative risk values is impacted by zero total events (Fig. 1). The inclusion of zero total events results in the majority of relative risk values falling between 1.02–1.11, with another 30% between 0.99–1.01 and 1.12–1.21 and few values outside the range 0.99–1.21. In comparison, excluding zero total events leads to a shift of the distribution towards larger values: the mode of the distribution of relative risk is between 1.22 and 1.31, with around 60% of the counties having relative risk values between 1.12 and 1.41. Overall, the inclusion of zero total events results in lower estimates of crash relative risk and a reduced range of values. Spatial patterns are also affected by the inclusion or exclusion of zero total events (Fig. 2). Exclusion of zero total events results in greater values of relative risk in most counties and much higher proportion of significant relative risk increases. Nearly 60% of the counties see a significant increase in crash relative risk when zero total events are excluded compared to 35% when zero total events are included. As the overall distribution of relative risk values showed, exclusion also results in a number of relative risk reductions; however, only one county in northeast Minnesota has a significant reduction in relative risk.

The inclusion or exclusion of zero total event pairs has a significant impact on the relative risk estimates. For zero total events, both A_i and B_i in Eq. (1) are zero, while C and D represent the sample sizes for the rainfall event and the dry control period. Zero total events can provide some information on the odds when C and D are known (Parzen et al., 2002). However, crash relative risk studies usually set C and D in an arbitrary manner due to the difficulty in collecting data on the number of total driving operations. If both A_i and B_i are zero and C and D are set to an arbitrary constant, then a zero total event pair would provide no information about the odds and should be excluded from the overall relative risk estimate. No mention of zero total events was found in the review of published studies on relative risk and it is possible that these studies reached the same conclusion on the elimination of zero total events; however, given the potential impact on relative risk estimates it is important to mention explicitly their exclusion, or the reasons for their inclusion. Subsequent analyses in this manuscript present relative risk with zero total events excluded.

3.2. Effects of temporal aggregation

Overall, relative risk estimates (excluding zero total events) at the daily and hourly temporal scale do not vary significantly for 75% of the study counties (Fig. 3). Nearly 95% of rural and 86% of rural interstate counties see no significant difference in relative risk when calculated at the daily versus hourly scale. In comparison, only 39% of urban

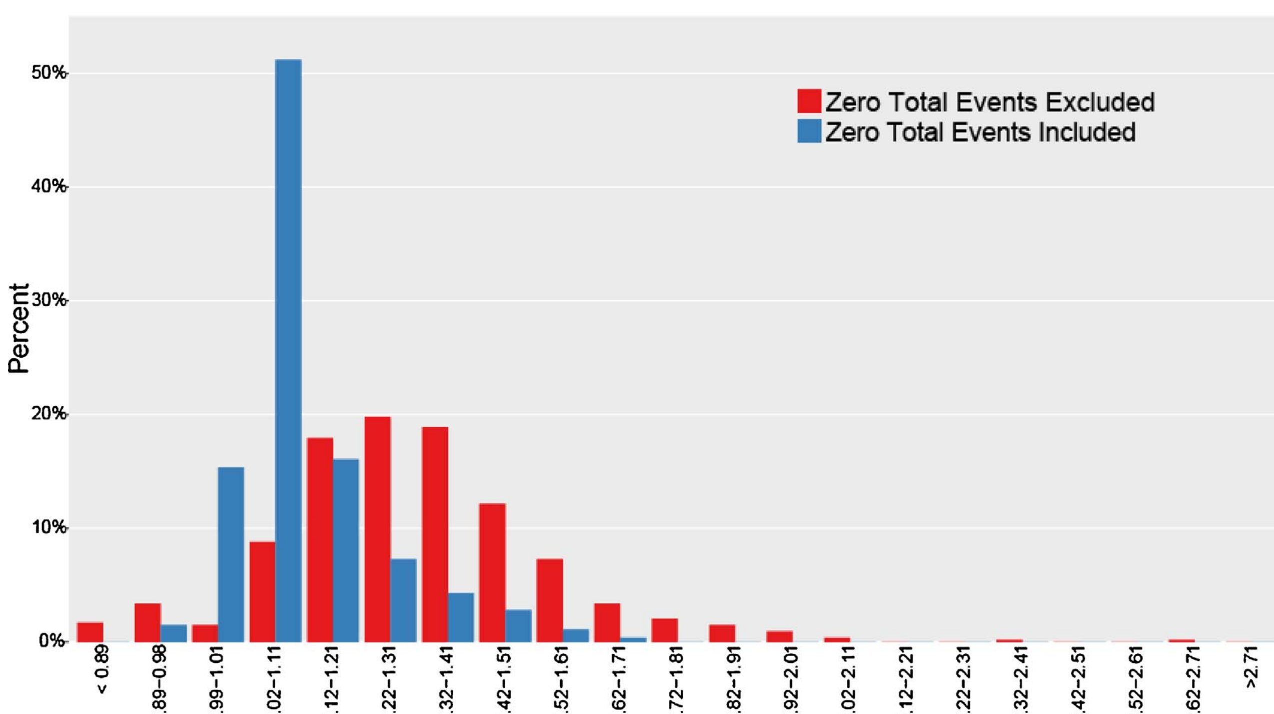

Fig. 1. Percentage of county level relative risk values in each category with zero total events excluded (red bars) and included (blue bars). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

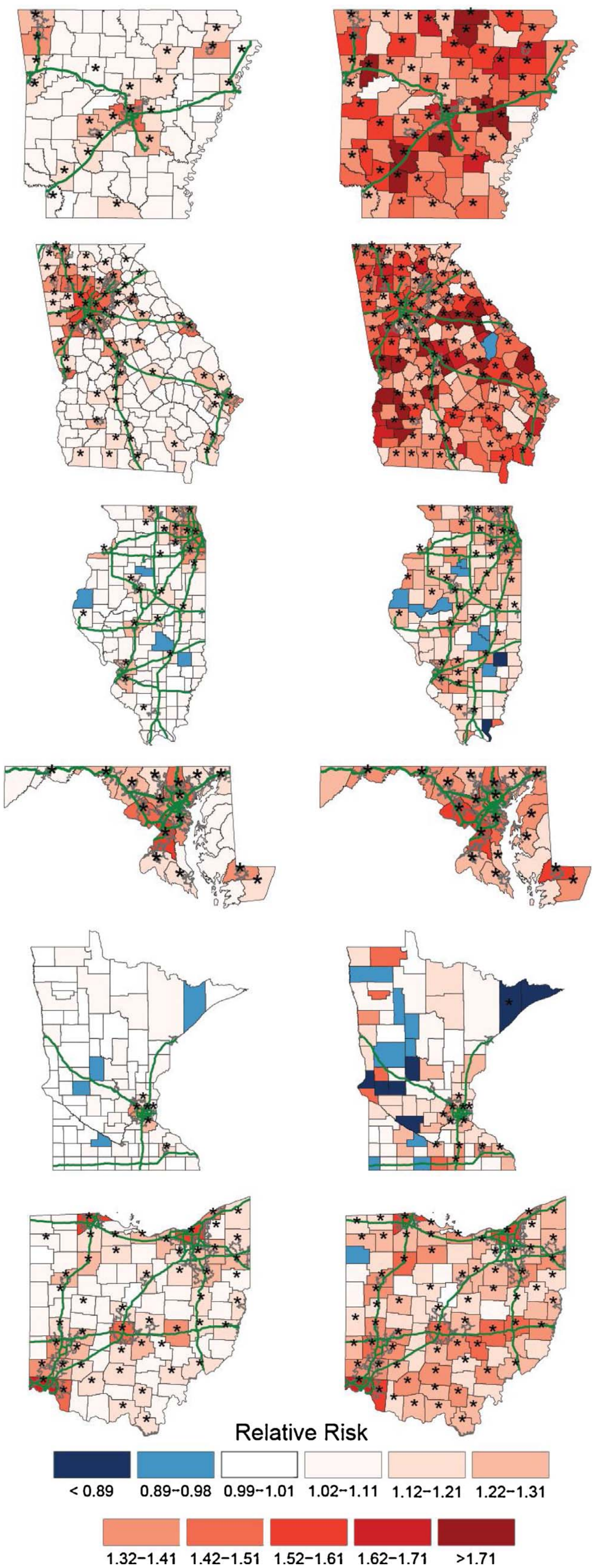

Fig. 2. County level relative risk of crash during rainfall based on hourly data for the six study states (from top to bottom) Arkansas, Georgia, Illinois, Maryland, Minnesota, and Ohio. The panels in the left (right) column are for the case with zero total events included (excluded). Counties that saw a relative risk different from 1.00 at the 5% level are marked with an asterisk. Interstate highways appear in green, and urban areas are outlined in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

counties have no significant difference in relative risk between daily and hourly scale. However, this does not mean that hourly data are necessarily required to estimate relative risk in urban areas. For 85% of urban counties, the use of six-hourly meteorological data would result in an estimate of relative risk that does not vary significantly from the estimate based on hourly data. Of the 15% of urban counties where the relative risk estimate based on the hourly data does vary significantly from a six-hourly estimate, the six-hourly estimate would be an average of 17% lower than the hourly estimate. If three-hourly meteorological data were available to estimate relative risk, only four urban counties would still see an estimate that varies significantly from the estimate based on hourly data: Gwinnett County, GA (Atlanta, GA area); Cook County, IL (Chicago); and Franklin (Columbus) and Hamilton (Cincinnati) counties in Ohio. While crash relative risk can be estimated in the majority of counties based upon daily data, urban counties where most crashes occur and that are of most interest in terms of relative risk during rainfall may exhibit significant differences in relative risk estimates when computed using daily and sub-daily data temporal resolution. It is important to note that our analysis is not designed to determine whether daily or sub-daily time steps should be used when analyzing relative risk, but rather to assess the differences in relative risk estimates due to temporal scale and determine if those differences are significant.

3.3. Estimates of relative risk from crash reports

Excluding zero total events, crash and injury relative risk are significantly higher when calculated using the standard approach as compared to relative risk calculated using rain in reports or wet pavement approaches (Fig. 4). For example, the crash relative risk is 39%, 16%, and 5% greater during rainfall as compared to dry conditions when calculated using the standard, wet pavement, and rain in reports methods respectively. This is not necessarily a surprising result given that the standard approach would presumably capture many more crashes during events. The difference in relative risk estimates among the three approaches narrows at higher hourly rainfall totals. The standard approach produces estimates of relative risk that are between 23% and 34% greater than the wet pavement and rain in reports methods, respectively, for all rainfall crashes; however, the difference between the standard approach and the wet pavement and rain in reports methods are only 16% and 23% for hours with at least 12.7 mm of rain. Relative risk estimates are higher when using the wet pavement approach than the rain-in-reports approach, a logical result given that wet pavement conditions are likely to persist for some time after rainfall has stopped. Similarly, the differences in relative risk estimates between the wet pavement and rainfall-in-reports methods are smaller for greater hourly precipitation, where there is a greater likelihood that rainfall occurred for a larger fraction of the hour.

Ultimately, our analysis is not designed to determine which approach provides the "best" estimates of relative risk, but instead to assess the change in relative risk that results from the different approaches. It is possible that using the information from crash reports could be better in some situations and the meteorological data could be better in others. For example, data from crash reports could be superior for short duration rainfall events that may not be detected by hourly meteorological observations, while the meteorological observations are not subject to potential errors in the coding of the weather or pavement conditions at the time of crash that could be present in crash report data. Overall, using the standard method will result in larger relative risk estimates; however, the difference between methods is less pronounced at higher precipitation thresholds. While significant differences exist in relative risk estimates between the wet pavement and rain-in-reports methods for all rainfall crashes, the difference is not significant at the highest precipitation threshold ($\geq$12.7 mm).

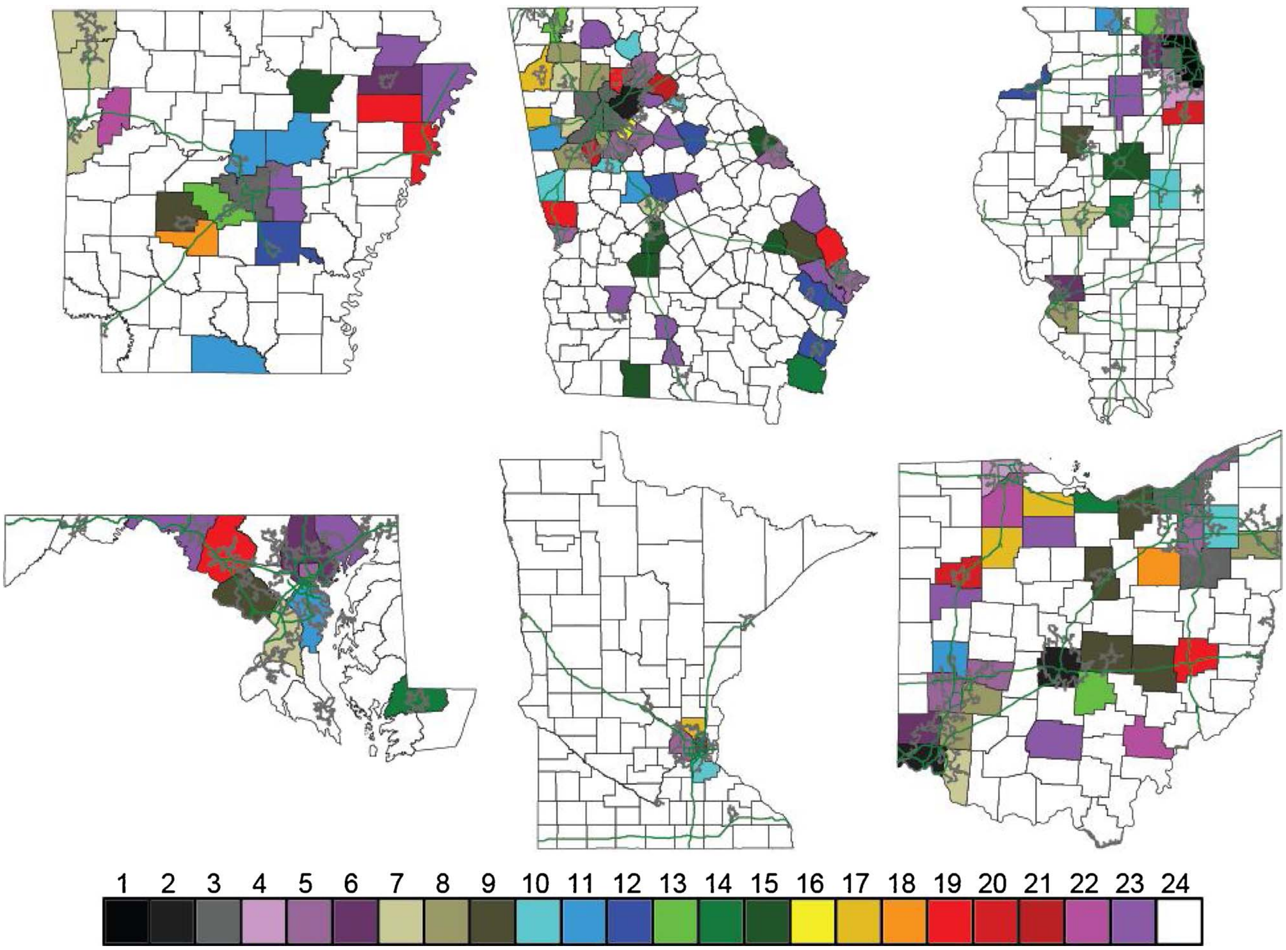

Fig. 3. Maximum temporal aggregation (in hours) to obtain an estimate of relative risk that is not statistically different (at the 5% level) from the one obtained at the hourly scale with zero total events excluded for the six study states of Arkansas, Georgia, Illinois (top row from left to right) and Maryland, Minnesota, and Ohio (bottom row left to right). Interstate highways appear in green, and urban areas are outlined in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

4. Conclusions

As noted by Jaroszweski and McNamara (2014), the methodological decisions made during analysis of crash relative risk due to adverse weather conditions have a significant impact on the estimates of relative risk obtained. By using a fine temporal and spatial scale of analysis, we assessed the effects of several methodological choices on relative risk estimates. Ultimately, we found that the methodological choices that were examined in this study do have a significant impact on relative risk estimates.

While many studies of crash risk have used the matched pair odds ratio method, none of them have discussed the impact and disposition of matched pairs where no crashes were observed during both the event and control — so called “zero total events.” Our analysis finds that the inclusion of zero total events has a significant impact on relative risk estimates, resulting in lower overall crash relative risk and a narrower distribution of values. Having excluded zero total events, we find that the temporal scale of analysis also has considerable effects on relative risk estimates. For over 75% of counties in the six study states, including 95% of rural and 86% of rural interstate counties, no statistically significant difference exists in the relative risk estimates calculated using hourly versus daily data. Urban areas are more sensitive to the temporal scale used, but there is no significant difference in relative risk estimated using six-hour data as compared to hourly data in 85% of urban counties. As expected, relative risk estimates (excluding zero total events) are lower when using the wet pavement and rainfall in reports approaches when compared to the standard method that assumes all crashes during the specified period are associated with rainfall. The difference in estimates is lowest when comparing the standard method to the wet pavement method and at higher hourly precipitation totals.

Based on our analysis, we recommend that studies of relative risk that follow the typical odds ratio methods and assign of an arbitrary value to C and D in Eq. 1 remove zero total events from their analysis as they do not contribute directly to the odds. We also recommend that authors clearly specify whether zero total events are included or excluded. Because the temporal scale of the analysis and the use of information from crash reports will greatly impact the subsequent relative risk estimates, we encourage authors to consider these impacts when designing their studies and discuss them when presenting their results. It is hoped that a more thorough understanding of the effect of methodological choices and their impacts will ultimately reduce their influence on relative risk estimates.

Acknowledgements

Alan Black and Gabriele Villarini acknowledge financial support by IIHR-Hydroscience & Engineering and the Iowa Flood Center. Gabriele Villarini acknowledges funding by the USACE Institute for Water Resources. Alan Black acknowledges support from the NOAA Regional Integrated Sciences and Assessments (RISA) Program from NOAA’s Climate Program Office under Cooperative Agreement NA13OAR4310183. The comments and suggestions by two anonymous reviewers are also gratefully acknowledged.

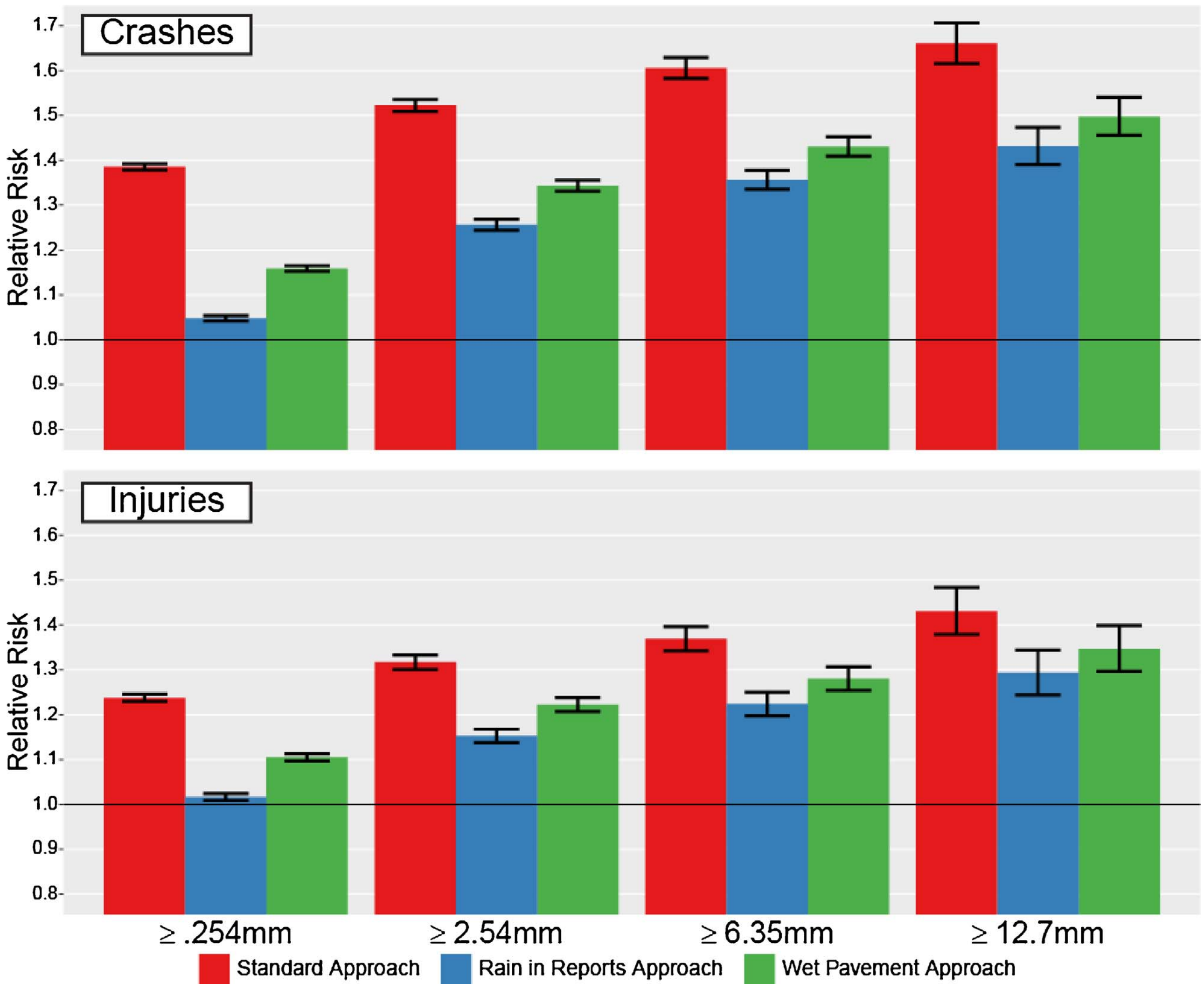

Fig. 4. Relative risk of crash (top) and injury (bottom) and 95% confidence intervals based on the standard approach (using hourly rainfall), rain in reports approach, and wet pavement approach for each precipitation threshold with zero total events removed. The black bars on each column represent the 95% confidence intervals of the estimated relative risk.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2018.01.023.

References

Abdel-Aty, M., Ekram, A., Huang, H., Choi, K., 2011. A study on crashes related to visibility obstruction due to fog and smoke. Accid. Anal. Prev. 43, 1730–1737.
Andrey, J., 2010. Long-term trends in weather-related crash risks. J. Transp. Geogr. 18, 247–258.
Andrey, J., Hambly, D., Mills, B., Afrin, S., 2013. Insights into driver adaptation to inclement weather in Canada. J. Transp. Geogr. 28, 192–203.
Andrey, J., Mills, B., Leahy, M., Suggett, J., 2003. Weather as a chronic hazard for road transportation in Canadian cities. Nat. Hazards 28, 319–343.
Andrey, J., Yagar, S., 1993. A temporal analysis of rain-related crash risk. Accid. Anal. Prev. 4, 465–472.
Bergel-Hayat, R., Debbarh, M., Antoniou, C., Yannis, G., 2013. Explaining the road accident risk: weather effects. Accid. Anal. Prev. 60, 456–465.
Black, A.W., Mote, T.L., 2015a. Characteristics of winter-precipitation-related transportation fatalities in the United States. Weather Climate Soc. 7, 133–145.
Black, A.W., Mote, T.L., 2015b. Effects of winter precipitation on automobile collisions, injuries, and fatalities in the United States. J. Transp. Geogr. 48, 165–175.
Black, A.W., Villarini, G., Mote, T.L., 2017. Effects of rainfall on vehicle crashes in six U.S. States. Weather Climate Soc. 9, 53–70.
Brodsky, H., Hakkert, A.S., 1988. Risk of a road accident in rainy weather. Accid. Anal. Prev. 20, 161–176.
Eisenberg, D., 2004. The mixed effect of precipitation on traffic crashes. Accid. Anal. Prev. 36, 637–647.
Elvik, R., Høye, A., Vaa, T., Sørensen, M., 2009. The Handbook of Road Safety Measures, 2nd Ed. Emerald Group Publishing, pp. 1124.
Elvik, R., 2006. Laws of accident causation. Accid. Anal. Prev. 38, 742–747.
Fleiss, J.L., Levin, B., Paik, M.C., 2003. Statistical Methods for Rates and Proportions –, third edition. John Wiley & Sons, Hoboken, New Jersey.
Gart, J.J., Zweifel, J.R., 1967. On the bias of various estimators of the logit and its variance with application to quantal bioassay. Biometrika 54, 181–187.
Haldane, J.B.S., 1955. The estimation and significance of the logarithm of a ratio frequencies. Ann. Hum. Genet. 20, 309–311.
Hambly, D., Andrey, J., Mills, B., Fletcher, C., 2013. Projected implications of climate change for road safety in Greater Vancouver. Can. Clim. Change 116, 613–629.
Jaroszweski, D., McNamara, T., 2014. The influence of rainfall on road accidents in urban areas: a weather radar approach. Travel Behav. Soc. 1, 15–21.
Johansson, Ö., Wanvik, P., Elvik, R., 2009. A new method for assessing the risk of accident associated with darkness. Accid. Anal. Prev. 41, 809–815.
Keay, K., Simmonds, I., 2005. The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. Accid. Anal. Prev. 37, 109–124.
Lin, Y., Mitchell, K.E., 2005. The NCEP stage II/IV hourly precipitation analyses: development and applications San Diego, CA, 9-13 January 2005, Paper 1.2. Preprints, 19th Conf. on Hydrology, American Meteorological Society.
Mills, B.N., Andrey, J., Hambly, D., 2011. Analysis of precipitation-related motor vehicle collision and injury risk using insurance and police record information for Winnipeg, Canada. J. Saf. Res. 42, 383–390.
Parzen, M., Lipsitz, S., Ibrahim, J., Klar, N., 2002. An estimate of the odds ratio that always exists. J. Comput. Graph. Stat. 11, 420–436.
Qiu, L., Nixon, W.A., 2008. Effects of adverse weather on traffic crashes. Transp. Res. Rec. 2055, 139–146.
Sherretz, L.A., Farhar, B.C., 1978. An analysis of the relationship between rainfall and the occurrence of traffic accidents. J. Appl. Meteor. 17, 711–715.
Sweeting, M.J., Sutton, A.J., Lambert, P.C., 2004. What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data. Stat. Med. 23, 1351–1375.
U.S. Census Bureau, 2015. 2010 Urban and Rural Classification. Accessed 7 March 2016. Available online at:. https://www.census.gov/geo/reference/ua/urban-rural-2010.html].