CHAPTER

7

Before–after studies in highway safety

7.1 Introduction

Before–after studies are frequently used for assessing the safety impacts of various interventions, such as the introduction of red-light cameras at signalized intersections or widening shoulder widths along rural two-lane highways. Because of its popularity, a lot of research has been devoted to studying either the methodologies associated with before–after studies (Hauer, 1997; Lord and Kuo, 2012; Li et al., 2019) or their applications for evaluating the safety effects of interventions (Persaud et al., 2001). The main hypothesis is that the before and after study has a better control on external factors as the same group of drivers, for example, are assumed to be using or traveling through the sites under study, which in return may more accurately capture the effects of the intervention (Hauer, 1997). Statistically speaking, compared with the cross-sectional study, the before–after study has lower within-subject variability (i.e., the variation associated with multiple measurements observed over time for one subject) (Kuo and Lord, 2013).

In general terms, the before–after study consists of predicting the anticipated number of crashes that would have occurred in the "after period," if the treatment(s) had not been implemented, and comparing the predicted value with the estimated number of crashes (usually the observed number of crashes) after the treatment was applied (note: the treatment can include one or more interventions, such as changing the lane and shoulder widths simultaneously). This is the basic approach that has been proposed by Hauer (1997) in his seminal textbook "Observational Before–After Studies in Road Safety.". This chapter builds on this

basic approach and expands on it by including other recent methods that have been proposed for conducting such studies.

This chapter first describes the two critical issues that can negatively influence this type of study. The next section presents the basic methods for conducting a before–after study without and with comparison groups. Then, the empirical Bayes (EB) and full Bayes (FB) methods are presented. Later, we cover other more recent methods, such as the naïve adjustment method, the before–after study using survival analysis, and the propensity score (PS) method. The chapter ends with a discussion about the sample size needed for conducting before–after studies.

7.2 Critical issues with before–after studies

This section is divided into two subsections. The first subsection describes the regression-to-the-mean (RTM) phenomenon, while the second subsection covers the site selection effects. Although these two are related, they produce distinct biases and affect before–after studies differently.

7.2.1 Regression-to-the-mean

As explained by other researchers (Abbess et al., 1981; Hauer and Persaud, 1983, 1984), the RTM dictates that when observations are characterized by very high (or low) values in a given time period and for a specific site (or several sites) (N_{before}), it is anticipated that observations occurring in a subsequent time period (N_{after}), are more likely to regress toward the long-term mean of a site ($\overline{N}$), as described in Fig. 7.1. The main assumption here is that each site remains unchanged over time (e.g., exposure, driving population). The RTM was first noted more than a century ago by Francis Galton (Stingler, 1997). In traffic safety, the RTM influences the estimated changes that were or would be made on the entities under analysis, as sites with large counts are usually selected for treatments (as described further below).

The RTM can be conceptualized mathematically using random variables in two time periods, labeled as 1 and 2, respectively (Copas, 1997; Lord and Kuo, 2012). Let us assume that Y_1 and Y_2 are two random variables with almost exactly the same distribution, but where the conditional expectation $E[Y_2|Y_1]$ is not equal to Y_1. It can be shown that the conditional expectation can be defined as a jointly normal distribution (Copas, 1997):

$$E[Y_2|Y_1] = \rho Y_1 + (1-\rho)\mu \tag{7.1}$$

where ρ is the correlation between Y_1 and Y_2, and μ is the common mean. When the correlation coefficient is equal to 1, no RTM exist as $E[Y_2|Y_1] = Y_1$. On the other hand, when the correlation coefficient is not equal to 1,

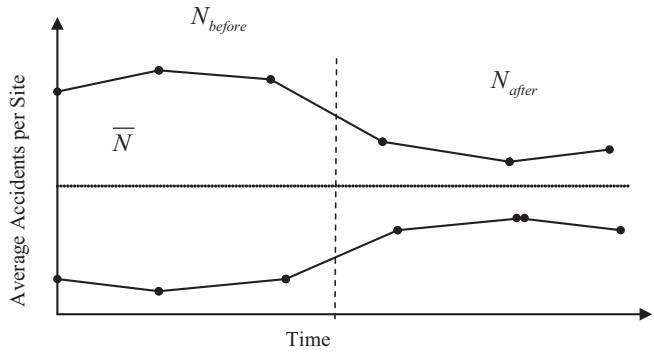

FIGURE 7.1 Representation of the regression-to-the-mean (Lord and Kuo, 2012).

RTM is observed in the data. Smaller values of ρ are associated with larger RTM effects because $E[Y_2|Y_1]$ is closer to μ and farther away from Y_1. Eq. (7.1) also shows that the magnitude of the RTM can be computed by taking the difference between $E[Y_2|Y_1]$ and Y_1 (Fig. 7.2). It is important to point out that the RTM only exists in the context of examining observations over time, such as those used for before-after studies. RTM does not exist in cross-sectional studies.

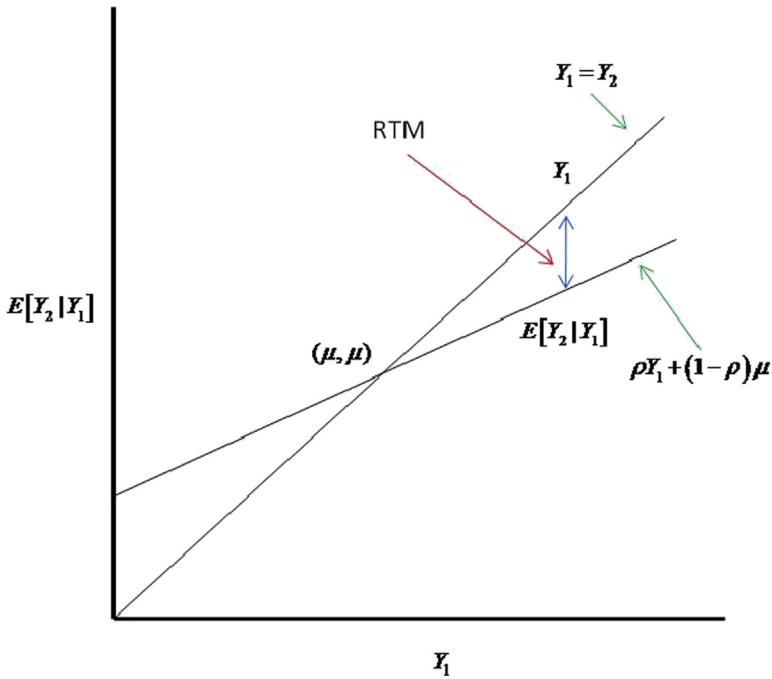

FIGURE 7.2 Relationship between $E[Y_2|Y_1]$ and Y_1 (Lord and Kuo, 2012).

7.2.2 Site selection bias

The general idea of site selection bias or effects is that setting entry criteria will transfer the original population distribution to a truncated sample distribution, resulting in changes in the characteristics of the new distribution such as the mean and variance of the data. Hence, ignoring these changes will create a biased estimator for the safety effectiveness (Fig. 7.3). The premise is that setting a higher entry criterion will cause higher site selection bias. This bias has been sporadically discussed in the literature (Hauer, 1980a,b; Abbess et al., 1981; Davis, 2000), but it has not been fully quantified in before–after studies until the work of Lord and Kuo (2012).

When selecting sites for treatment, transportation agencies can select these sites based on a required minimum of the number of crashes. For instance, Warrant 7 of the Manual on Uniform Traffic Control Device (MUTCD) that is used for justifying the installation of traffic signals based solely on safety states that a site must experience five or more failure-to-yield crashes within a 12-month period (FHWA, 2009). When such traffic

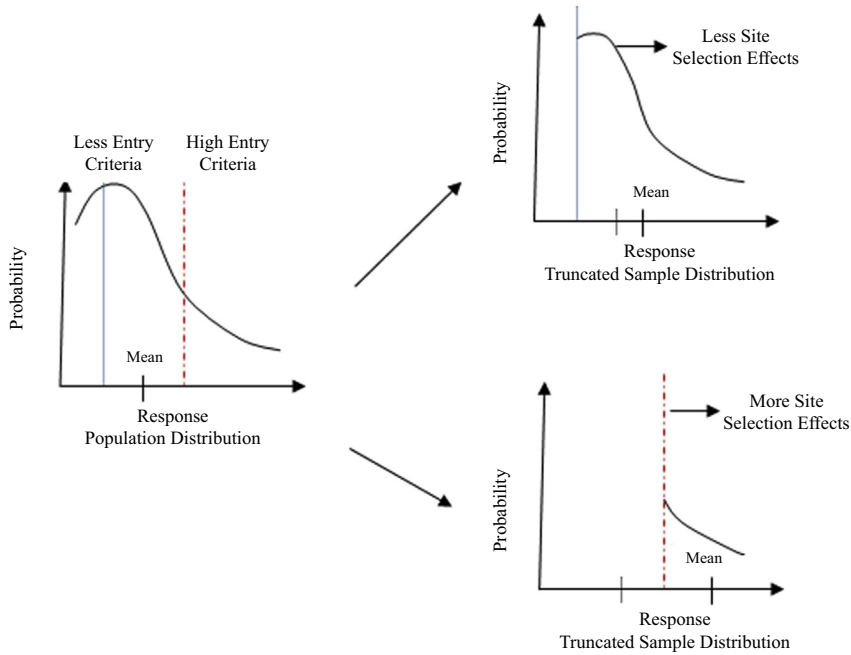

FIGURE 7.3 The population distribution for complete and truncated samples (Lord and Kuo, 2012).

signals are evaluated to address a safety problem, the sample should theoretically be extracted from sites that met Warrant 7 of the MUTCD. Setting a minimum entry criterion is not limited to practitioners, as researchers have also used entry criteria for evaluating safety projects (Miranda-Moreno et al., 2009). Furthermore, in many other cases, the selection effect exists but is never explicitly spelled out, because the safety evaluation of a treatment often includes sites that are greater than 0, even if the sites selected for treatment are not chosen specifically for safety reasons. When the entry criterion is small, say when all the sites experience at least one crash, there is a bias. The reader is referred to Lord and Kuo (2012) and Kuo and Lord (2013) for more detailed information about the effects of the RTM and site selection biases on before–after studies.

7.3 Basic methods

This section presents the two basic methods for conducting before–after studies.

7.3.1 Simple before–after study

As described earlier, the goal for conducting before–after studies is to compare the predicted value if the treatment or change was not implemented with the estimated number of crashes in the after period (usually the actual number of observed crashes) for the target crash (e.g., right-angle, run-of-the-road, all crashes) and their severity (Fatal injuries, Incapacitating injuries, Nonincapacitating injuries, Possible injuries, Property Damage Only or a combination thereof). The assessment can be accomplished by taking the difference or the ratio (referred to as "the index" from now on) (Hauer, 1997):

$$\delta = \pi - \lambda \tag{7.2a}$$

$$\theta = \lambda/\pi \tag{7.2b}$$

where δ is the reduction in the expected number of target crashes; θ is the index or the ratio between the estimated and expected number of target crashes; π is the expected number of target crashes of a specific entity in the after period if the treatment had not been implemented (the number is a predicted value); and, λ is the estimated number of target crashes of a specific entity in the after period (as described above, the estimated value

is usually the number of observed crashes in the after period). If δ is negative, this implies that the number of crashes increased in the after period.

For a simple or naïve before–after study, the "before" period is used as the predicted value for the after period. The predicted value for the after period can be adjusted for the time differences between the before and after periods (r_d), as well as the linear or nonlinear relationship between crashes and traffic flow (r_{tf}) (note: the relationship is the same for both time periods). The predicted value is adjusted as follows (Hauer, 1997):

$$\pi = r_d r_{tf} \kappa \tag{7.3}$$

where r_d is the ratio in terms of duration (i.e., months, years, etc.) between the after period and the before period; r_{tf} is the ratio in terms of traffic flow between the after period and the before period; and, κ is the unadjusted number of crashes in the before period. The ratio of the traffic flow is given by the following equation (Hauer, 1997):

$$r_{tf} = \left(F_A/F_B\right)^{\beta} \tag{7.4}$$

where F_A is the traffic flow entering or traveling through an entity during the after period (vehicles/day); F_B is the traffic flow entering or traveling through an entity during the before period (vehicles/day); and, β is the relationship between crashes and traffic flow (usually estimated via a regression model, as discussed in Chapter 3—*Crash-Frequency Modeling*: $\mu = F^{\beta}$). If β is equal to 1, then the relationship is linear.

Eqs. (7.2a)–(7.4) are used for each site individually. If several sites are analyzed as a group, then the analyst needs to take the summation. Hence, for Eq. (7.5) for example, we get

$$\theta = \sum_{i=1}^{n} \lambda_i \Big/ \sum_{i=1}^{n} \pi_i \tag{7.5}$$

If the ratios are different for each site, then Eq. 7.5 becomes

$$\theta = \sum_{i=1}^{n} \lambda_i \Big/ \sum_{i=1}^{n} r_{di} r_{tfi} \kappa_i \tag{7.6}$$

To simplify the analysis, the summation function for the rest of the equations is not included here, but the values are assumed to be a summation of all the sites or observations in the group in the next two equations.

The variance or uncertainty associated with the estimated values above are given as follows:

$$Var(\lambda) = \lambda \tag{7.7a}$$

$$Var(\pi) = \pi, \text{ if a simple before} - \text{after study is conducted} \tag{7.7b}$$

$$Var(\pi) = r_d^2 \kappa, \text{ if only the ratio of the time periods is used to adjust for the predicted value} \tag{7.7c}$$

$$Var(\pi) = r_d^2 \times \left(r_{tf}^2 \kappa + \kappa^2 v_{tf}\right), \text{ if the predicted value is adjusted using both ratios.} \tag{7.7d}$$

where $v_{tf} = Var(r_{tf}) = r_{tf}^2 \beta^2 \times \left[cv_{\text{After}}^2 + cv_{\text{Before}}^2\right]$ and cv = the percent coefficient of variation in traffic flow for the before and after time periods. In practice, the percent coefficient of variation can be very difficult to obtain. Hence, if it is not available, values between 0.10 and 0.20 could be used in Eq. (7.7d). It is recommended to conduct a sensitivity analysis to estimate how sensitive the cv is for different values.

Hauer (1997) recommends adjusting the index θ for the small sample size bias (i.e., sample of crashes fewer than 500 crashes or observations in the before period):

$$\theta = \frac{\lambda}{\pi[1 + Var(\pi)/\pi^2]} \tag{7.8}$$

The variance of the estimated reduction and index are given as follows:

$$Var(\delta) = Var(\pi) + Var(\lambda) \tag{7.10}$$

$$Var(\theta) \cong \theta^2 \left[\left(Var(\lambda)/\lambda^2\right) + \left(Var(\pi)/\pi^2\right)\right] \tag{7.11}$$

If the sample size in the before period is smaller than 500 crashes, the variance of θ is adjusted as follows:

$$Var(\theta) \cong \frac{\theta^2 \left[\left(Var(\lambda)/\lambda^2\right) + \left(Var(\pi)/\pi^2\right)\right]}{\left[1 + Var(\pi)/\pi^2\right]^2} \tag{7.12}$$

It should be pointed out that $Var\{\pi\}/\pi^2$ will tend towards 0 when the sample size becomes greater than 500.

Exercise 7.1

Using the Before–After Dataset, conduct a before–after study using the naïve method. Assume a treatment was installed at the beginning of Year 6 (i.e., after period is 3 years). The data were collected over 8 years for 15 sites.

First, calculate the ratio r_d

$$r_d = 3/5 = 0.6$$

Estimate π, λ, $Var(\pi)$, $Var(\lambda)$

Site ID	r_d	π	λ	$Var(\pi)$	$Var(\lambda)$
1	0.6	7.2	5	4.3	5
2	0.6	9	9	5.4	9
3	0.6	9.6	5	5.8	5
4	0.6	9.6	5	5.8	5
5	0.6	15.6	9	9.4	9
6	0.6	8.4	5	5.0	5
7	0.6	15	12	9.0	12
8	0.6	11.4	9	6.8	9
9	0.6	11.4	16	6.8	16
10	0.6	10.8	14	6.5	14
11	0.6	17.4	8	10.4	8
12	0.6	15.6	12	9.4	12
13	0.6	3.6	11	2.2	11
14	0.6	8.4	8	5.0	8
15	0.6	18.6	12	11.2	12
Sum		171.6	140	103.0	140

Calculate δ

$$\delta = \pi - \lambda$$
$$\delta = 171.6 - 140$$
$$\delta = 31.6$$

There is a reduction of 31.6 in the expected number of crashes.

Calculate θ (adjust for a sample size below 500)

Exercise 7.1 (*cont'd*)

$$\theta = \frac{\lambda}{\pi[1 + Var(\pi)/\pi^2]}$$

$$\theta = \frac{140}{171.6[1 + 103.0/171.6^2]}$$

$$\theta = 0.82$$

There is a reduction of 18% in the expected number of crashes.

Calculate $Var(\delta)$ and $SD(\delta)$

$$Var(\delta) = Var(\pi) + Var(\lambda)$$

$$Var(\delta) = 103.0 + 140$$

$$Var(\delta) = 243$$

$$SD(\delta) = \sqrt{\delta}$$

$$SD(\delta) = \sqrt{243.0}$$

$$SD(\delta) = 15.6$$

The reduction is $31.6 \pm 1.96 \times 15.6$, which is statistically significant at the 5% level.

Calculate $Var(\theta)$ and $SD(\theta)$

$$Var(\theta) \cong \frac{\theta^2\left[\left(Var(\lambda)/\lambda^2\right) + \left(Var(\pi)/\pi^2\right)\right]}{\left[1 + Var(\pi)/\pi^2\right]^2}$$

$$Var(\theta) \cong \frac{0.82^2\left[\left(140/140^2\right) + \left(103.0/171.6^2\right)\right]}{\left[1 + 103.0/171.6^2\right]^2}$$

$$Var(\theta) \cong 0.007$$

$$SD(\theta) = \sqrt{0.007}$$

$$SD(\theta) = 0.084$$

The reduction is $0.82 \pm 1.96 \times 0.084$, which is statistically significant at the 5% level.

Exercise 7.2

Redo Exercise 7.1, but only include sites that experienced three or more crashes in the before period. Use the last 3 years before and the last 3 years after (i.e., $r_d = 1$).

First, estimate π, λ, $Var(\pi)$, $Var(\lambda)$

Site ID	r_d	π	λ	$Var(\pi)$	$Var(\lambda)$
1					
2	1	10.0	9	10.0	9
3	1	11.0	5	11.0	5
4					
5					
6					
7	1	13.0	12	13.0	12
8					
9	1	11.0	16	11.0	16
10					
11	1	20.0	8	20.0	8
12	1	20.0	12	20.0	12
13					
14					
15	1	22.0	12	22.0	12
Sum		107.0	74	107	74

Calculate δ

$$\delta = 33.0$$

There is a reduction of 33.0 in the expected number of crashes.

Calculate θ (adjust for a sample size below 500)

$$\theta = 0.69$$

There is a reduction of 31% in the expected number of crashes, which is much larger than the reduction observed in Exercise 7.1. This shows the effects of the site selection bias.

Calculate $Var(\delta)$ and $SD(\delta)$

$$Var(\delta) = 181.0$$
$$SD(\delta) = 13.5$$

The reduction is $33.0 \pm 1.96 \times 13.5$, which is statistically significant at the 5% level.

Exercise 7.2 (*cont'd*)

Calculate $Var(\theta)$ and $SD(\theta)$

$$Var(\theta) \cong 0.011$$
$$SD(\theta) = 0.103$$

The reduction is $0.69 \pm 1.96 \times 0.103$, which is statistically significant at the 5% level.

7.3.2 Before–after study with comparison groups

In the previous section, the estimated reduction and index are based entirely on the data collected at sites where one or more treatments/changes are implemented. The estimate is not adjusted for factors that could influence crash risk on a regional level or very large scale. For example, weather patterns, economic conditions (i.e., growth, recessions, etc.) (as discussed in Chapter 1—*Introduction*), and crash reportability thresholds should influence all observations in a similar manner, whether they are part of the treatment group or not. Those are called external factors. To account for these changes, the before–after study can be supplemented by incorporating sites or observations that are not the subject of a treatment, referred to as control, reference, or comparison sites; for the rest of the chapter, we will use the terminology comparison group. This section, hence, describes how these sites can be used in before–after studies.

Including comparison sites as part of the before–after analysis relies on two assumptions (Hauer, 1997):

(1) The external factors that affect safety have changed from the before to the after period in the same manner on both groups.
(2) This change in the extarnal factors affects the treatment and the comparison groups the same way.

Furthermore, the sites that are part of the comparison group should ideally have similar characteristics as those in the treatment group. These include variables such as the sample mean, sample variance and range of traffic flow, lane and should widths, presence of crosswalks at intersections. Note that, in practice, it may be very difficult to match all these characteristics between both groups. Nonetheless, the comparison group can be useful to capture changes in large-scaled factors.

TABLE 7.1 Notation definition for treatment and comparison groups. (Hauer, 1997)

Period	Treatment group	Comparison group
Before	κ	μ
After	λ	ν

The ratio between the before and after periods for the comparison group is defined as r_c. The notations are defined in Table 7.1. Recall that the variables are summations of the sites in each group.

Using the notation described in Table 7.1, the ratios can be calculated for the treatment and comparison group, respectively, as follows[1]:

$$r_t = \sum \pi \Big/ \sum \kappa \tag{7.13}$$

$$r_c = \sum \nu \Big/ \sum \mu \tag{7.14}$$

Similar to what was described earlier, we have $\pi = r_t \kappa$. Given the assumptions mentioned earlier, the hope is that $r_c = r_t$, which means $\pi = r_t \kappa = r_c \kappa$; the term π is used in Eq. 7.13 rather than λ since it denotes that a treatment was applied. Note that the ratio r_c in Eq. 7.14 is an approximation, but should be valid for most datasets (see (Hauer, 1997)). For this relationship to work, the time periods (before and after) should be the same for the treatment group and comparison group. If they are not, both groups could be divided in subsamples (or eventually for each individual site) by matching observations in both groups that have the same time periods before and after. Then, conduct the analysis for each subsample and continue or finalize the analysis using the concept described in Eq. 7.5 (the summation of the subsamples or individual sites). Note that the adjustment for changes in traffic flow may not be needed since it is indirectly captured via the comparison group with the assumption that the relationship between crash risk and flow is the same for both groups and time periods. If they are not and traffic flows are available, (Hauer, 1997) has proposed different options for including r_{tf} and readers are referred to his textbook for more details.

The assessment can again be accomplished the same way as before (Hauer, 1997):

$$\delta = \pi - \lambda \tag{7.15}$$

$$\theta = \lambda / \pi \tag{7.16}$$

[1] For the rest of this section, we will remove the summation variable Σ to simplify the description.

Exercise 7.3

Redo Exercise 7.1, but consider the 25 sites that were included in the comparison group. Use 5 years before and 3 years after.

(Note: the value of the following parameters shows the summation over 15 sites.)

First, calculate r_c

$$r_c = \frac{\nu}{\mu}$$

$$r_c = \frac{259}{405}$$

$$r_c = 0.64$$

Estimate π, λ, $Var(\lambda)$

Site ID	r_c	π	λ	$Var(\lambda)$
1	0.64	7.7	5	5
2	0.64	9.6	9	9
3	0.64	10.2	5	5
4	0.64	10.2	5	5
5	0.64	16.6	9	9
6	0.64	9.0	5	5
7	0.64	16.0	12	12
8	0.64	12.2	9	9
9	0.64	12.2	16	16
10	0.64	11.5	14	14
11	0.64	18.5	8	8
12	0.64	16.6	12	12
13	0.64	3.8	11	11
14	0.64	9.0	8	8
15	0.64	19.8	12	12
Sum		182.9	140	140

Calculate $Var(\pi)$

$$Var(r_t)/r_t^2 \approx 1/\mu + 1/\nu + Var(\omega)$$

$$Var(r_t)/r_t^2 \approx 1/405 + 1/259 + 0.001$$

$$Var(r_t)/r_t^2 \approx 0.0073$$

$$Var(\pi) = \pi^2\left[1/\kappa + Var(r_t)/r_t^2\right]$$

$$Var(\pi) = 182.9^2[1/286 + 0.0073]$$

$$Var(\pi) = 362.2$$

Calculate δ

$$\delta = 42.9$$

continued

Exercise 7.3 (*cont'd*)

There is a reduction of 42.9 in the expected number of crashes.

Calculate θ (adjust for a sample size below 500)

$$\theta = 0.77$$

There is a reduction of 23% in the expected number of crashes. This value is larger than the reduction observed in Exercise 7.1. The data from the comparison group shows a small reduction in the number of crashes as well, which explains why θ decreased from 82% to 77%.

Calculate $Var(\delta)$ and $SD(\delta)$

$$Var(\delta) = 502.2$$
$$SD(\delta) = 22.4$$

The reduction is $42.9 \pm 1.96 \times 22.4$, which is not statistically significant at the 5% level (almost at the boundary).

Calculate $Var(\theta)$ and $SD(\theta)$

$$Var(\theta) \cong 0.010$$
$$SD(\theta) = 0.101$$

The reduction is $0.77 \pm 1.96 \times 0.101$, which is statistically significant at the 5% level.

The only difference is that the variance of the predicted value (π) is calculated as follows:

$$Var(\pi) = \pi^2 \left[1 / \kappa + Var(r_t) / r_t^2\right] \tag{7.17}$$

where,

$$Var(r_t) / r_t^2 \approx 1/\mu + 1/\nu + Var(\omega) \tag{7.18}$$

$Var(\omega)$ is the variance of the odds ratio $\omega = r_c/r_t$.

The variance of the odds ratio can be calculated as described in Hauer (1997). However, the authors of this textbook suggest using $Var\{\omega\} = 0.001$, which is a good approximation for most cases.

$Var(\delta)$ and $Var(\theta)$ are calculated the same way as for the simple method.

7.4 Bayesian methods

There are two general methods that employ the Bayesian theory. Those are known as the FB method and the EB method. The EB method was initially developed as an approximation of the Bayes method, as the latter method requires a multidimensional integration of the total prior function

of the Bayesian (bottom of Eq. 7.19), which was first described in Chapter 2—*Fundamentals and Data Collection*:

$$f(\mu|y) = \frac{p(y|\mu)f(\mu)}{\int_a^b p(y|\mu)f(\mu)d\mu} \tag{7.19}$$

However, over the last 30 years or so, Bayesian models or techniques have become much more popular with the application of the Markov Chain Monte Carlo (MCMC) (and other estimation techniques) and with the incredible advancement in computing power. Although the FB method can now be efficiently used for conducting before–after studies, the EB method remains very popular as the latter method is relatively easy to implement or use. It is now a well-established method in the AASHTO's Highway Safety Manual (AASHTO, 2010).

7.4.1 Empirical Bayes method

The EB method has been described extensively by others, such as Hauer (1997) and Persaud and Lyon (2007). The idea is to use two pieces of information, one coming from the site investigated and one coming from a population of sites that are assumed to have the same basic characteristics (e.g., 4-legged urban signalized intersections, 4-lane divided rural arterials). It is highly recommended that the population of sites in the comparison has similar mean, range and variance values for key variables, such as traffic flow, lane and shoulder widths, as those for the sites in the treatment group. This assumption is addressed further below.

Furthermore, the EB method was initially developed as a function of the negative binomial (NB) distribution/regression model as this distribution, based on the mixture of the Poisson and Gamma distributions, has a closed form and can be easily manipulated mathematically[2] (Abbess et al., 1981; Hauer, 1986). Hence, the following description is based on that relationship.

The estimate of the EB method is given by Eq. (7.20):

$$\mu_{i\_EB} = \gamma_i \mu_i + (1 - \gamma_i) y_i \tag{7.20}$$

where μ_{i_EB} is the long-term mean estimated using the EB method for site i; μ_i is the long-term mean based on sites with similar characteristics, which can be estimated using the method of moment or a regression model; y_i is the number of crashes for site i; γ_i is the weight factor specific

[2] Although the EB method was initially developed using the NB distribution or model, researchers have recently introduced the EB for other models, such as the finite mixture and the Sichel models (see Zou et al., 2013; Ash et al., 2017).

to each site i. It should be pointed out that Eq. 7.20 is basically used for obtaining a better estimate of the long-term mean of a site.

The weight factor is calculated by the following equation:

$$\gamma_i = \frac{1}{1 + \frac{\mu_i}{\phi}} \tag{7.21}$$

with ϕ being the inverse dispersion parameter of the NB model. As discussed in Chapter 3—*Crash-Frequency Modeling*, the dispersion parameter can be made dependent on the characteristics of the each site (ϕ_i). If several years of data are used in the EB analysis, the term μ_i (in crashes/year) in Eq. 7.21 is multiplied by the number of years n and y_i in Eq. 7.20 is the total number of observed crashes for the n years under study. The curious reader will notice that Eq. 7.1 and Eq. 7.20 are very similar, where the weight factor is equivalent to 1-ρ. In the latter equation, if the weight factor is equal to zero (perfect correlation or $\rho=1$), then there is no RTM as the EB estimate, μ_{i_EB} (i.e. the long-term mean), is equal to the number of observed crashes.

The variance of the EB is calculated by Eq. (7.22):

$$Var\{\mu_{i\_EB}\} = (1 - \gamma_i)\mu_{i\_EB} \tag{7.22}$$

The following are the steps needed for conducting a before–after study with the EB (Persaud et al., 2001).

7.4.1.1 Step 1—collect data for the treatment and comparison groups

Step 1 consists of collecting and assembling data for conducting the before–after study. The same variables are collected for both groups. The reader is referred to Chapter 2—*Fundamentals and Data Collection* for information about how to collect and assemble safety data.

It should be pointed out that enough data need to be collected for developing reliable NB regression models. As discussed in Lord (2006), the sample size is governed by the characteristics of the sample mean of the data. The recommended sample size (number of site *i*s) was presented in Table 6.3 in Chapter 6—*Cross-Sectional and Panel Studies in Safety*.

7.4.1.2 Step 2—develop a regression model from the comparison group

In this step, the safety analyst needs to develop an NB model that will be used for estimating the long-term mean of site i based on the characteristics of the data collected for the comparison group. The NB model is estimated using the following relationship:

$$\mu_i = \exp(\mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i) \tag{7.23}$$

where μ_i is the long-term mean for site i (in crashes/year); $\boldsymbol{\beta}_i$ is a vector of estimable parameters; $\mathbf{x}'$ is a vector of explanatory variables; and, $\exp(\varepsilon_i)$

is a gamma-distributed error term with mean 1 and variance α or $\alpha = 1/\phi$. The term α refers to the dispersion of the dispersion parameter of the NB model and ϕ is the inverse dispersion parameter. This variable is needed for calculating the weight factor of the EB method described earlier. The variance of the estimate is calculated using the following equation:

$$Var\{\mu_i\} = \frac{\mu_i^2}{\phi} \tag{7.24}$$

7.4.1.3 Step 3—estimate the EB for the before period

Using the NB model developed in Step 2, estimate the long-term mean based on the EB method for each site i. The EB for the before period is estimated as follows:

$$\mu_{i\_EB} = \frac{(\phi + y_{i\_b})}{\left(\phi/\mu_{i\_b} + t_{i\_b}\right)} \tag{7.25}$$

where μ_{i_EB} is the EB estimate for site i (in crashes/year); μ_{i_b} is the regression output of the NB model for site i for the before period; y_{i_b} is the number of crashes in the before period for site; t_{i_b} is the number of months or years in the before period; and ϕ is the inverse dispersion parameter.

The variance of the EB estimate is given as follows:

$$Var\{\mu_{i\_EB}\} = (1-\gamma)\mu_{i\_EB} \tag{7.26}$$

7.4.1.4 Step 4—calculate r_{tf}

Similar to the simple before-after study, there is a need to adjust for the differences in traffic flow for each site i. The ratio is calculated using this relationship:

$$r_{i\_tf} = f(A)/f(B) \tag{7.27}$$

where, $f(A) = \mu_{i\_a} = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ and $f(B) = \mu_{i\_b} = \exp(\mathbf{x}_i'\boldsymbol{\beta})$.

In this step, the analyst needs to apply the NB model for each site for both the before and after periods.

7.4.1.5 Step 5—estimate the predicted value for the after period

The predicted value is estimated by multiplying the EB estimate for the before period with the ratio r_{tf} and the number of months or years (same time unit) in the after period:

$$\pi_i = r_{i\_tf} \times t_a \times \mu_{i\_EB} \tag{7.28}$$

Sum all the sites together:

$$\pi = \sum_{i=1}^{n} \pi_i \tag{7.29}$$

7.4.1.6 Step 6—calculate the estimated value for the after period

Calculate the estimate value for the after period, which is the number of crashes for each site (λ_i).

Sum all the sites together:

$$\lambda = \sum_{i=1}^{n} \lambda_i \tag{7.30}$$

7.4.1.7 Step 7—calculate the variance for the predicted and estimated values

The variance for the predicted is calculated as follows:

$$Var(\pi_i) = \frac{\mu_{i\_EB} \times (r_{i\_tf} \times t_{i\_a})^2}{\left(\phi/\mu_i + t_{i\_b}\right)} \tag{7.31}$$

where all the terms are as they were defined earlier. The variance of the estimated value defined as follows:

$$Var(\lambda_i) = \lambda_i \tag{7.32}$$

Sum all the sites together:

$$Var(\pi) = \sum_{i=1}^{n} \pi_i \tag{7.33a}$$

$$Var(\lambda) = \sum_{i=1}^{n} \lambda_i \tag{7.33b}$$

7.4.1.8 Step 8—calculate the difference and index

This step is the same one as described earlier:

$$\delta = \pi - \lambda \tag{7.34}$$

$$\theta = \lambda/\pi \tag{7.35}$$

Note that the adjustment to correct for the small sample size bias also applies here.

7.4.1.9 Step 9—calculate the variance for the difference and index

In addition, the variance for the difference and index is calculated the same way as for the naïve and comparison group methods:

$$Var\{\delta\} = Var\{\pi\} + Var\{\lambda\} \tag{7.36}$$

$$Var\{\theta\} \cong \theta^2 \left[\left(Var\{\lambda\}/\lambda^2 \right) + \left(Var\{\pi\}/\pi^2 \right) \right] \tag{7.37}$$

Similar to Step 8, the variance needs to be adjusted for the small sample size bias of the treatment group.

7.4.1.9.1 Caution with the EB method

Although the EB method has now become a common or very popular method for estimating the safety effects of interventions in before–after studies, it has recently been documented that the estimate can still be biased if the characteristics of the treatment and comparison sites differ significantly (when the comparison sites are used as input for developing NB or other types of crash-frequency models). Davis (2000) first raised this issue, but the bias was fully described and quantified by Lord and Kuo (2012) who used simulated data and found that the estimate will be biased if the sample mean and variance of the treatment and comparison group datasets are different. In before–after studies, the difference with these characteristics is expected to be observed as the sites selected for treatment are usually selected for a reason (i.e., site selection). Hence, if a site experiences a large number of crashes, it is more likely to be included in the treatment group rather than in the comparison group. How this bias affects the EB estimate was further confirmed by Wood and Donnell (2017) and Li et al. (2019), who used simulated and observed data.

7.4.2 Bayes method

As described earlier, with the advancements in computing power and the application of the MCMC simulation, developing Bayes or FB models is now relatively easy to perform. The main advantage of using the Bayes method is that the treatment and comparison groups can be combined into one dataset for the before and after periods, and the effect of the treatment estimated accordingly. Furthermore, the EB method assumes that the covariate effect on crashes is known with certainty, whereas the Bayes method assumes that the covariates are represented by a distribution (the posterior estimate to be exact). Using the formulation described in Chapter 3—*Crash-Frequency Modeling*, the general model, assuming that each site i follows a Poisson count, is defined in this manner:

$$u_{it} = \exp(\mathbf{x}_{it}\boldsymbol{\beta}_{it} + \varepsilon_i) \tag{7.38}$$

where μ_{it} is the mean of site i and time t; $\mathbf{x}_{it}$ is a vector of covariates for site i and time t; $\boldsymbol{\beta}_{it}$ is a vector of covariates for site i and time t; and, $\exp(\varepsilon_i)$ is the error that can follow a gamma or lognormal distribution.

In the context of a before–after study, the vector of covariates can be defined using the modeling characteristics proposed by Park et al. (2010). Let the elements of the vector $\mathbf{x}_{it} = (x_{1it}, \ldots, x_{6it}, x_{7i}, \ldots, x_{ki})$ be $x_{1it} = \ln(AADT_{it})$ (if a nonlinear relationship is used); $x_{2it} = T_i$; $x_{3it} = t$; $x_{4it} = (t - t_{oi})I[t > t_{0i}]$; $x_{5it} = T_i t$; $x_{6it} = T_i(t - t_{oi})I[t > t_{0i}]$; and, $x_{7i}, \ldots, x_{ki}$ are

roadway or traffic-related characteristics (e.g., lane width, percentage of trucks, etc.)

Where $T_i = 1$ if the ith is a treatment site and zero otherwise; t is the tth in the study period; t_{0i} is the year in which the countermeasure or treatment was installed (for a site in a comparison group, this is defined as the same year as that for the treatment group); and, $\mathbf{I}[t > t_{0i}] = 1$ if t belongs to the after period or zero otherwise.

Eq. (7.38) can be rewritten as follows (Park et al., 2010):

$$\mu_{it} = \exp\begin{pmatrix} \beta_0 + \beta_1 \ln AADT_{it} + \beta_2 T_i + \beta_3 t + \beta_4 (t - t_{0i})\mathbf{I}[t > t_{0i}] \\ +\beta_5 T_i t + \beta_6 T_i (t - t_{0i})\mathbf{I}[t > t_{0i}] + \beta_7 x_{7i} + \ldots + \beta_k x_{ki} \end{pmatrix} \quad (7.39)$$

The parameters β_3 to β_6 play a role in the estimation of the before–after study. It can be shown that Eq. (7.39) can be reformulated under the framework of a before–after study (Park et al., 2010), such that:

$$\mu_{it,\text{Control},B} = \exp(\beta_0 + \beta_1 \ln AADT_{it} + \beta_3 t + \beta_7 x_{7i} + \cdots + \beta_k x_{ki}) \quad (7.40a)$$

$$\mu_{it,\text{Control},A} = \exp((\beta_0 - \beta_4 t_{0i}) + \beta_1 \ln AADT_{it} + (\beta_3 + \beta_4) t + \beta_7 x_{7i} + \cdots + \beta_k x_{ki}) \quad (7.40b)$$

$$\mu_{it,\text{Treatment},B} = \exp((\beta_0 + \beta_2) + \beta_1 \ln AADT_{it} + (\beta_3 + \beta_5) t + \beta_7 x_{7i} + \cdots + \beta_k x_{ki}) \quad (7.40c)$$

$$\mu_{it,\text{Treatment},A} = \exp\begin{pmatrix} \{\beta_0 + \beta_2 - (\beta_4 + \beta_6) t_{0i}\} + \beta_1 \ln AADT_{it} \\ +(\beta_3 + \beta_4 + \beta_5 + \beta_6) t + \beta_7 x_{7i} + \cdots + \beta_k x_{ki} \end{pmatrix} \quad (7.40d)$$

As discussed in Chapter 2—*Fundamentals and Data Collection*, the coefficients of the FB model require to make an assumption about the hyper-priors and they need to be estimated using the MCMC (or an alternative method). The model above directly handles the RTM by comparing the mean before and after for the treated sites (Pawlovich et al., 2006; Li et al., 2008). However, as discussed earlier, the site selection bias could still negatively influence this method (Lord and Kuo, 2012; Xie et al., 2019), as the input data similarities between the treatment and comparison groups apply here as well.

After specifying the hyper-parameters and estimating the posterior values of the model's coefficients, obtain the posterior distribution of the estimated crashes for each group: before and after estimates for the treatment, $\mu_{TB} = \sum_{it} \mu_{it,\text{Treatment},B}$ and $\mu_{TA} = \sum_{it} \mu_{it,\text{Treatment},A}$; and, the before and after estimates for the comparison group, $\mu_{CB} = \sum_{it} \mu_{it,\text{Control},B}$ and $\mu_{CA} = \sum_{it} \mu_{it,\text{Control},A}$, respectively.

Using the formulation described in Section 7.2.2, the index and the difference are calculated as follows:

7.4.2.1 Step 1—calculate R_c

Estimate the posterior distribution, point estimate and standard deviation of R_c for the comparison group:

$$R_c = \frac{\mu_{CA}}{\mu_{CB}} \tag{7.41}$$

7.4.2.2 Step 2—predict π

Estimate the posterior distribution, point estimate and standard deviation of the predicted crashes:

$$\pi = \mu_{TB} \times R_c \tag{7.42}$$

7.4.2.3 Step 3—estimate θ

Estimate the posterior distribution, point estimate, and standard deviation of the index:

$$\theta = \frac{\mu_{TA}}{\pi} \tag{7.43}$$

7.4.2.4 Step 4—estimate δ

Estimate the posterior distribution, point estimate, and standard deviation of the difference:

$$\delta = \pi - \mu_{TA} \tag{7.44}$$

7.4.2.5 Step 5—determine the significance of θ and δ

Estimate the 2.5-, 5-, and 10-percentile from the posterior distribution of the index and the difference. Then, compare the values with the nominal condition if the expected reduction (or increase) is statistically significant (e.g., if the x-percentile of θ includes the value 1 or not).

The advantage with the full Bayes method over the EB method is that the uncertainty can be directly estimated in Step 5, whereas for the EB method, the variance is estimated using an extra step (Step 7 in the EB method).

7.5 Adjusting for site selection bias

As discussed earlier, site selection bias occurs when an entry criterion is used for selecting observations that will be included in the before–after

study (most before–after studies have such bias even if it is not explicit). For count data, this gives rise to a truncated NB distribution, which was described in Fig. 7.3. Kuo and Lord (2017) have proposed a simple method for reducing the bias with the index θ and the difference δ when a comparison group is not available. Recall that the EB and the FB methods can also be negatively affected by the site selection bias. The method does not eliminate the site selection bias but can reduce it up to about one half of the entire bias. The RTM and the site selection effect can only be fully removed if the characteristics of the comparison group, such as the sample mean and sample variance, is exactly the same as for the treatment group, which is very difficult to collect in practice, as discussed above.

Kuo and Lord (2017) have proposed the following equations, initially based on the theoretical work described in Lord and Kuo (2012), for reducing the site selection bias for the index and the difference:

$$\theta_{adj} = \theta_{naive}\left[1 + \frac{C+1 \Big/ \left(1 + \frac{P(N>C+1)}{P(N=C+1)}\right)}{\Lambda_1 + \left[\frac{C+1}{(\Lambda_1 \times \alpha_{naive}+1)^{-1}\left(1+\frac{P(N>C+1)}{P(N=C+1)}\right)}\right]}\right] \tag{7.45}$$

$$\delta_{adj} = \delta_{naive} - \left[\frac{(C+1)\times(\delta_{naive}\alpha_{naive}-1)}{1+\frac{P(N>C+1)}{P(N=C+1)}}\right] \tag{7.46}$$

where,

θ_{adj} is the adjusted index,

δ_{naive} is the adjusted difference.

$\theta_{naive} = \frac{\Lambda_2}{\Lambda_1} = \frac{\frac{1}{n}\frac{1}{t}\sum_{i=1}^{n}\sum_{j=1}^{t}N_{ij2}}{\frac{1}{n}\frac{1}{t}\sum_{i=1}^{n}\sum_{j=1}^{t}N_{ij1}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t}N_{ij2}}{\sum_{i=1}^{n}\sum_{j=1}^{t}N_{ij1}}$, the safety index.

$\delta_{naive} = \Lambda_2 - \Lambda_1 = \frac{1}{n}\frac{1}{t}\left(\sum_{i=1}^{n}\sum_{j=1}^{t}N_{ij2} - \sum_{i=1}^{n}\sum_{j=1}^{t}N_{ij1}\right)$, the difference between the mean values,

C is The entry criterion ($C \geq 0$; e.g., when $C = 0$, the entry criterion is equal to $Y = 1$).

n is the sample size.

$\alpha_{naive} = \frac{\left[\frac{E\left((N_{i2}-\Lambda_2)^2|N_{i1}>C\right)}{E(N_{i2}|N_{i1}>C)}-1\right]}{E(N_{i2}|N_{i1}>C)}$, the dispersion parameter.

Λ_{i1} is the mean response rate for site i in the before period, $i = 1, \ldots, n$.

Λ_{i2} is the mean response rate for site i in the after period, $i = 1, \ldots, n$.

N_{ij1} is the observed response for site i in j year (in the before period) for count data, N_{ij1}>C.

N_{ij2} is the observed response for site i in j year (in the after period) for count data. Here, $j = t = 1$ for calculation convenience purposes.

To simplify the calculations documented in Eqs. (7.45) and (7.46), Kuo and Lord (2017) have developed a spreadsheet that can be used for reducing the bias for the index and difference (see Appendix A of the paper). The user just needs to input the data into the spreadsheet. The spreadsheet can be found at https://ceprofs.civil.tamu.edu/dlord/Papers/Adjusted%20table.xlsx.

7.5.1 Example application for estimating θ_{adj}

In this no-treatment before–after evaluation example (Kuo and Lord, 2017), it was assumed that a dummy treatment was applied to 917 sites (which had at least one recorded crash in 2008) in College Station, Texas, on December 31, 2008. In this case, the year 2008 was defined as the before period, while the year 2009 was defined as the after period. Because there was no such applied treatment, the index should be close to 1. In reality, the actual index of this dummy treatment (based on all of the crash data available, without a positive criterion selection) was 0.95 or a 5% reduction.

For this example, the entry criterion was set to five crashes per year, $Y > C = 4$. From 917 sites, 121 sites were identified as having a crash record equal to 5 crashes or above per year in the before period, $Y_1 = [5, 5, \ldots, 31, 32]$, while the corresponding crashes in the after period were $Y_2 = [5, 4, \ldots, 41, 32]$.

The adjusted index, θ_{adj}, can be calculated as follows.

7.5.1.1 Step 1—calculate the naïve estimate

$$\theta_{naive} = \frac{\Lambda_2}{\Lambda_1} = \frac{1/121\,(5+5+\mathrm{K}+31+32)}{1/121\,(5+4+\mathrm{K}+41+32)} = \frac{983}{1217} = 0.808 \tag{7.47}$$

Recall that the time period is 1 year before and 1 year after.

7.5.1.2 Step 2—estimate the value of the variables inside Eq. (7.45)

$$\Lambda_1 = \frac{1217}{121} = 10.06 \tag{7.48}$$

$$\alpha_{naive} = \frac{\left[\dfrac{E\left((N_{i2}-\Lambda_2)^2 | N_{i1} > C\right)}{E(N_{i2}|N_{i1} > C)} - 1\right]}{E(N_{i2}|N_{i1} > C)} = \frac{(60.8/8.12) - 1}{8.12} = 0.799 \tag{7.49}$$

$$\frac{P(N > C+1)}{P(N = C+1)} = \frac{0.600}{0.060} = 9.97 \tag{7.50}$$

The probability and cumulative probability, Eq. (7.50), were estimated using the "dnbinom" and "dnbinom" functions in the R software program. The spreadsheet has a build-in functions for calculating these two probabilities.

7.5.1.3 Step 3—calculate the adjusted safety index (Eq. 7.45)

$$\theta_{adj} = \theta_{naive}\left[1 + \frac{C+1\Big/\left(1+\frac{P(N>C+1)}{P(N=C+1)}\right)}{\Lambda_1 + \left[\frac{C+1}{(\Lambda_1 \times \widehat{\alpha}_{naive}+1)^{-1}\left(1+\frac{P(N>C+1)}{P(N=C+1)}\right)}\right]}\right] \tag{7.51a}$$

$$\theta_{adj} = 0.808\left[1 + \frac{5/(1+9.97)}{10.06 + \left(\frac{5}{(10.06\times 0.799+1)^{-1}(1+9.97)}\right)}\right] \tag{7.51b}$$

$$\theta_{adj} = 0.834$$

In this case, the θ_{adj} does not need to be adjusted for the small sample size (>500 observations in the before period).

7.6 Propensity score matching method

Recently, some safety analysts have proposed using the propensity score matching (PSM) method for estimating the safety effect of treatments or roadway features (see, e.g., Davis, 2000; Aul and Davis, 2006; Wood and Donnell, 2017; Wood et al., 2015). This method consists of matching pairs of treated and untreated units by their propensity scores—a scalar summary of a set of multiple potentially confounding covariates (e.g., traffic and roadway characteristics). This systematic matching intends to control for the confounding effect of influential covariates, thus allowing the estimation of the treatment effect as the mean difference between pairs of treated and untreated units; latter is referred to as control units. Overall, the method tries to mimic randomized experiments by eliminating or reducing the site selection bias (discussed earlier). Technically, the PSM is not a method that would be classified as a before–after study, but is a method that compares sites

where a treatment was implemented with sites that were not (i.e., comparison group). It is included here as the PSM has often been compared to the EB method (Wood et al., 2015).

The propensity score, $P_i(T=1|\mathbf{x}_i)$, denotes the probability that a subject is assigned to the treated group $(T=1)$ given its characteristics (Rosenbaum and Rubin, 1983). The covariate distribution of $\mathbf{x}_i$ may be different across the treated group and the comparison group. However, conditional on the propensity score, the covariate distribution should be similar between the two groups (Rosenbaum and Rubin, 1983).

For the simplest method, the propensity scores can be estimated using a logit model using data collected from both groups:

$$P_i(T=1|\mathbf{x}_i)=\frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})} \tag{7.52}$$

where $\mathbf{x}_i$ is a vector of variables for site i and $\boldsymbol{\beta}$ is a vector of coefficients. Once the logit model is estimated, the analyst applies the model to each group to determine the conditional probability of a facility receiving a treatment or countermeasure given the covariates and the outcomes. The matching is accomplished by examining sites in each group where the probability are within an interval P_i+c of each other, where c is the caliper distance specified for matching (Sasidharan and Donnell, 2013). The matching can be done 1:1, meaning one site for each group are grouped together, or 1:n, where more than one site from the comparison group can be match for each site from the treatment group.

After the matching is completed, the balance of the coefficients for the propensity scores, here the logit regression, between the treatment and comparison groups need to be examined. Sasidharan and Donnell (2013) suggest the following equation to estimate the potential bias for each coefficient:

$$SB=\frac{100\left(\overline{x}_{jT}-\overline{x}_{jC}\right)}{\sqrt{\left(S_{jT}^2+S_{jC}^2\right)/2}} \tag{7.53}$$

where SB is the standardized bias, $\overline{x}_{jT}$ is the sample mean for coefficient j of the treatment group, $\overline{x}_{jC}$ is the sample mean for coefficient j for the comparison group, S_{jT}^2 is the sample variance for coefficient j for the treatment group, and S_{jC}^2 is the sample variance for coefficient j for the comparison group. For the coefficients that fall below a predetermined SB boundary (say below 5%), the coefficients are kept in forfurther analysis, as described in the following.

At the end of this step, the analyst can either estimate the treatment effect by conducting the summation of the subgroup for the treatment and comparison groups:

$$\delta = \frac{1}{n_2}\sum_i y_{iC} - \frac{1}{n_1}\sum_i y_{iT} \tag{7.54}$$

where δ is the treatment effect (a positive number refers to a positive gain), y_{iT} the number of crashes at site i for the treatment group, y_{iC} the number of crashes at site i for the comparison group, n_1 is the number of observations in the treatment group, and n_2 is the number of observations in the comparison group.

It is possible to estimate a crash-frequency model using the subgroup data and the coefficients that were identified using Eq. 7.55:

$$\mu_i = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right) = \exp\left(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i\right) \tag{7.55}$$

The characteristics of these models are described in Chapter 3—*Crash-Frequency Modeling*. The coefficients of the models can be used for estimating crash modification factors (CMFs).

It should be pointed out that, although propensity scoring methods can minimize the site selection bias, the variables need to be specifically matched, which is very difficult to do with observational studies (see Winkelmayer and Kurth, 2004; Austin, 2011; Biondi-Zoccai et al., 2011; Lim et al., 2014). As explained earlier, variables in the treatment and comparison groups can be very different. Recently, a Bayesian version of the PSM method was evaluated, but offered very little improvements over the traditional PSM method (Li and Donnell, 2020). It is possible, however, that if informative priors are used, the Bayesian version could be a better alternative to the likelihood-based PSM method.

7.7 Before–after study using survival analysis

So far, all the methods that have been described to assess the safety effect of treatments or changes in the design or operational characteristics of sites have used "raw" crash data. Recently, some researchers have proposed using the survival analysis framework as a method for conducting a before–after study (Xie et al., 2019). The advantage of the survival analysis framework is that it does not need to rely on comparison groups, which could negatively influence the EB and FB methods if the sample size is too small or the characteristics of the comparison group are different from the treatment group.

Under a survival analysis framework, the value of interest is the time between crashes and not the crashes themselves. This is called as the crash

hazard. If the number of crashes is assumed to be Poisson distributed over time, then the crash hazard is assumed to be exponentially distributed and can be defined as follows:

$$p(k,t|\lambda) = \frac{(\lambda t)}{k!}\exp(-\lambda t) \tag{7.56}$$

where p is the probability of observing k crashes during a time interval t; and, λ is the number of crashes per unit of time. It can be shown that the instantaneous crash risk at t can be estimated using the following hazard function (Allison, 2012; Xie et al., 2019):

$$H(t|\lambda) = \frac{f(t|\lambda)}{S(t|\lambda)} = \lambda \tag{7.57}$$

where $f(t|\lambda) = \lambda\exp(-\lambda t)$ is the PDF of the exponential distribution, as related to Eq. 7.56; and, $S(t|\lambda) = \exp(-\lambda t)$ is the survival function, which is the crash-free probability during interval t. Eq. 7.57 shows that the crash occurrence is the crash hazard λ, which is independent of time t. In the context of before–after study, the crash hazard λ becomes the dependent variable rather than crash count for the previous methods described earlier.

As discussed by Xie et al. (2019), censoring of the data (some observations) is needed as the time periods in before–after studies are finite or fixed and the time at which crashes occur may not always be known, especially at the aggregated level (the usual input data in the models).

Fig. 7.4 illustrates this concept, which shows when crashes occurred during the before and after periods, listed as T_{before} and T_{after}, respectively. The terms t_1, t_m, t_{m+1}, and t_K are right-censored as the time between crashes is cut-off by the boundaries of the before–after study and cannot be measured.

Xie et al. (2019) have proposed the Bayesian estimating method for conducting a before–after study using a survival analysis, as it combines prior distributions with the likelihood function that is derived from the

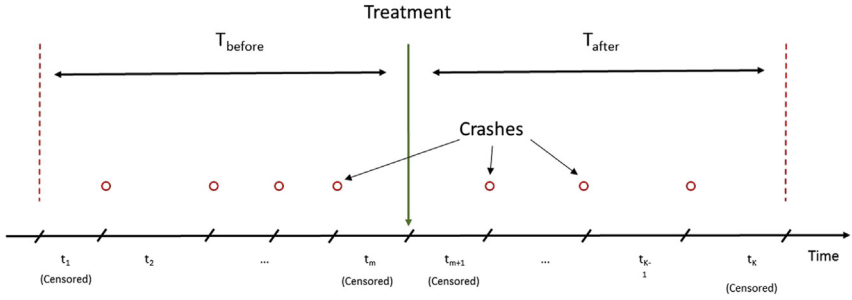

FIGURE 7.4 Censored data in before–after studies (Xie et al., 2019).

observed data. The likelihood function can be defined as follows (Ibrahim et al., 2001, pp. xiv, 479):

$$\begin{aligned} L(\boldsymbol{\beta}|K,\mathbf{t},\mathbf{x},\boldsymbol{\nu}) &= \prod_{k=1}^{K} f(t_k|\lambda_k)^{\nu_k} S(t_k|\lambda_k)^{(1-\nu_k)} \\ &= \prod \left[\exp\left(\mathbf{x}_k'\boldsymbol{\beta}\right)\exp\left(-t_k\exp\left(\mathbf{x}_k'\boldsymbol{\beta}\right)\right)\right]^{\nu_k} \\ &\quad \times \left[\exp\left(-t_k\exp\left(\mathbf{x}_k'\boldsymbol{\beta}\right)\right)\right]^{(1-\nu_k)} \\ &= \exp\left\{\sum_{k=1}^{K}\nu_k\mathbf{x}_k'\boldsymbol{\beta}\right\}\exp\left\{-\sum_{k=1}^{K}t_k\exp\left(\mathbf{x}_k'\boldsymbol{\beta}\right)\right\} \end{aligned}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_K)'$ is a vector of the sequence of time between successive crashes and is assumed to follow an exponential distribution; $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_K)'$ is a vector describing the censoring indicators, where $\nu_k = 0$ indicates t_k is right censored and $\nu_k = 1$ indicates t_k is a regular observation; and, $\mathbf{x}'$ and $\boldsymbol{\beta}$ are as previously described.

Similar to the FB method described earlier (Section 7.4.2), the treatment effect can be modeled directly in the Bayesian model using data collected for the treatment group:

$$\lambda_{il} = \exp\left(\beta_0 + \sum_{p=1}^{P}\beta_p x_{pil} + \beta_T \text{Treatment}_{il} + \varepsilon_i\right) \tag{7.59}$$

where λ_{il} is the crash hazard for site i and time interval l; p is the number of parameters in the model; β_0, β_p, and β_T are coefficients to be estimated (posterior estimate); Treatment_{il} is a dummy variable where 0 is used before treatment and 1 after; and, ε_i is the error term to capture the heterogeneity, such as the gamma or lognormal distribution. It can be shown that if the gamma distribution is used, the hazard function can revert back to the NB crash-frequency model.

The index is estimated as follows:

$$\theta = \exp(\beta_T) \tag{7.60}$$

Xie et al. (2019) showed that, using simulation, the before–after study with survival analysis performed better than the EB method when the comparison group's characteristics are very different from the treatment group.

7.8 Sample size calculations

To ensure that a treatment or change in an operational or design characteristics was (or will be) successfully implemented, one has to

collect enough data in the before and after time periods. This is an important task that should be considered whether it will be part of a future analysis (i.e., starting to collect the before data in the present time) or as part of a retroactive analysis (i.e., both the before and after time periods are in the past). If not enough data are collected, the safety analyst will not be able to make a proper assessment, given the study objective, such as the direction of the inference (increase or decrease in crash risk) or a value for the index or difference that may not be statistically significant for a given confidence level. If this happens, this could lead to a waste of funds and/or worse additional loss of lives and increases in injuries.

In simple terms, the sample size calculations are based on the inference statistics of the normal distribution:

$$\begin{aligned} P(\mu - 1\cdot\sigma(\mu) \le \mu \le \mu + 1\cdot\sigma(\mu)) &\approx 65\% \\ P(\mu - 2\cdot\sigma(\mu) \le \mu \le \mu + 2\cdot\sigma(\mu)) &\approx 95\% \\ P(\mu - 3\cdot\sigma(\mu) \le \mu \le \mu + 3\cdot\sigma(\mu)) &\approx 99\% \end{aligned} \tag{7.61}$$

Eq. (7.61) shows that confidence intervals vary from 65% to approximately 99%. The next three sections explain the important characteristics and methods for calculating the required sample size.

7.8.1 Factor influencing sample size calculations

Determining the appropriate (or minimum) sample size is governed by five factors. There are described as follows:

Variance of the variable being studied: This refers to examining the variation observed in the data, which will influence the index θ and difference δ. The variance is the expectation of the squared deviation of a random variable from its mean value. The standard deviation (SD) is the square root of the variance, while the standard error (SE) measures how variable the mean is from one sample to another. The standard error is

$$SE = \frac{SD}{\sqrt{N}} \tag{7.62}$$

In general, the greater the variance, the more data need to be collected. Although not used for before–after studies, Shirazi et al. (2016) describes in greater details data size requirements as a function of the variance and sample mean values.

Magnitude of the effects: This refers to the anticipated reduction in expected crashes. Before starting the data collection procedure, the analyst needs to estimate, either based on previous studies or personal experience, the expected reduction (or increase) that a treatment or a change in operational or design characteristic is expected to be observed. Greater is

the difference, fewer observations are needed. For example, if an analyst needs to appropriately detect a 5% reduction in crashes ($\theta = 0.95$) versus a 50% reduction ($\theta = 0.50$), the analyst will need to collect much more crashes for the former value than for the latter one.

Significance level: The significance level tells the analyst how likely it is that an observed difference is due to chance when the true difference is 0:

$$\begin{aligned} H_0 &: \theta_1 = \theta_2 \\ H_1 &: \theta_1 > \theta_2 \end{aligned} \tag{7.63}$$

Table 7.2 illustrates the four different situations used in hypothesis testing. The 2 cells with the correct decision indicate that H_0 is true and was appropriately not rejected for a given level of significance or H_0 was false and was appropriately rejected for a given level of power. The Type I error (also known as or aka false positive) is defined as the probability of rejecting the null hypothesis when it is actually true. The null hypothesis can usually be defined as "no difference in safety between the before and after periods." The Type II error (aka false negative) is defined as the probability of failing to reject the null hypothesis when it is actually false. The latter error is discussed further below.

In the context of Table 7.2, the significance level refers to the probability of rejecting the null hypothesis if it is true. Usually, three probabilities are used for evaluating the significance level: 10% ($p \leq 0.10$), 5% ($p \leq 0.05$), or 1% ($p \leq 0.01$) (note: it does not mean that the effect does not exist if the p-value is larger than the accepted threshold. It simply means that there is a greater probability that the observed change was caused by chance. The p-value is dependent on the sample size, with larger sample size leading to smaller p-values). The smaller is the probability, the more observations are needed.

Statistical power: The power is the probability that it will correctly lead to the rejection of a false null hypothesis. One can think of power as the probability of detecting a true effect. There are two different aspects about the power analysis. One is to calculate the necessary sample size for a

TABLE 7.2 Matrix describing the hypothesis testing.

	Do not reject H_0	**Reject H_0**
H_0 is true	Correct decision $1 - \alpha$: Confidence level	Type I error α: Significance level
H_0 is false	Type II error β	Correct decision $1 - \beta$: Power of a test

specified power. The other aspect is to calculate the power for a given specific sample size.

Generally, a test with a power greater than 0.8 (or $\beta \leq 0.2$) is considered statistically powerful. Increasing both the level of significance and power will increase the number of observations needed to conduct a before–after study, often significantly. The statistical power has a large effect on the sample size, as discussed further below. Although some researchers have recommended not using the statistical power (see Hauer, 2008), it may be useful to calculate if resources are available, as discussed next.

Cost of collecting data: As any data collection process, collecting crash and other related data can be a very expensive endeavor (see Chapter 2—*Fundamentals and Data Collection*). It is expensive as it often involves on-site (i.e., site visits) and on-line (e.g., Google Earth and Streetview) data collection procedures, as well as various tasks involving data reduction and assembling a wide variety of datasets into a common one that is ready to be used for conducting safety analyses. Usually, collecting data involves a large team of midlevel to junior workers (e.g., graduate students). Data collection often accounts between one-third and a half of the total expenditures for most research or highway safety-related engineering projects. In an ideal world, one should be able to collect a very large dataset that could be used for accurately estimating the safety effects, especially for very small differences. However, the selection of the sample size will always be a balancing act between the factors described above and the resources available to collect the necessary data.

7.8.2 Sample size estimation using known crash counts for both time periods

The sample size can be calculated by examining the difference in the expected number of crashes per year between the before and after periods. With this approach, the analyst must have access to information about the crash characteristics for both time periods of the sites under study.

First, let us define x_b and x_a as the number of crashes in the before and after periods, and t_b and t_a as the before and after time periods (say in years). Then, let $\mu_b = x_b/t_b$ and $\mu_a = x_a/t_a$ be defined as the number of crashes per unit of time (i.e., Poisson mean over time). Using the methodology proposed by Hauer (2008), one can calculate or examine the sample size based on this relationship $d > 0$, where $d = \mu_b - \mu_a$. In other words, the analyst examines if the difference in means is statistically

significant (say at a 1%, 5%, or 10%-level) or is significantly larger than 0. To do this, the analyst can use the following equation:

$$\frac{d}{SE(d)} = \frac{\mu_b - \mu_a}{\sqrt{(x_b/t_b^2 + x_a/t_a^2)}} = Z_{\alpha/2} \tag{7.64}$$

where, $Z_{\alpha/2=0.005} = 2.575$, $Z_{0.025} = 1.960$ or $Z_{0.05} = 1.645$. As explained earlier, the analyst needs to estimate the potential reduction in the number of crashes following a treatment or changes in operational or design characteristics. If the statistical power is included in the analysis, Eq. (7.64) becomes the following:

$$\frac{d}{SE(d)} = \frac{\mu_b - \mu_a}{\sqrt{(x_b/t_b^2 + x_a/t_a^2)}} = Z_{\alpha/2} + Z_{\beta} \tag{7.65}$$

One can see that if the power is included, the data requirement becomes larger. Table 7.3 can be used for obtaining values for the right-hand side of Eq. (7.65) for different combinations of significance and power.

The two equations above could be used for evaluating the significance level and/or study power after the data are collected. However, preferably, they should be used before the data collection process is enacted.

TABLE 7.3 Combination of significance and power (Kelsey et al., 1986).

Significance (α)	Power ($1 - \beta$)	$Z_{\alpha/2} + Z_{\beta}$
0.01 ($Z_{\alpha/2=0.005} = 2.575$)	0.80	3.417
	0.90	3.857
	0.95	4.221
	0.99	4.902
0.05 ($Z_{0.025} = 1.960$)	0.80	2.802
	0.90	3.241
	0.95	3.605
	0.99	4.286
0.10 ($Z_{0.05} = 1.645$)	0.80	2.802
	0.90	3.241
	0.95	3.605
	0.99	4.286

Exercise 7.4

A treatment is planned to be introduced at signalized intersections of a small city. On average, about 100 crashes per year occurred for 4 years during the before period (or 400 crashes in total). For the after period, about 90 crashes per year are expected to occur within the next 2 years. To be confident at 5% significance level, what would be the expected reduction with the installation of treatment?

First, select Eq. (7.64) as the study power is not required:

$$\frac{d}{SE(d)} = \frac{\mu_b - \mu_a}{\sqrt{(x_b/t_b^2 + x_a/t_a^2)}} = Z_{\alpha/2}$$

Enter the known values into the equation:

$$\frac{100 - \theta \times 90}{\sqrt{((100 \times 4)/4^2 + (90 \times 2)/2^2)}} = 1.96$$

where $\mu_b = x_b/t_b = (4 \times 100)/4 = 100$ and $\mu_a = x_a/t_a = (2 \times 90)/2 = 90$. Solve for θ

$$\begin{aligned} 100 - \theta \times 90 &= 1.96 \times \sqrt{(400/16 + 180/4)} \\ 100 - \theta \times 90 &= 1.96 \times \sqrt{70.0} \\ -\theta \times 90 &= 16.40 - 100 = -83.60 \\ \theta &= 85.47/90 = 0.93 \end{aligned}$$

The treatment is expected to reduce the number of crashes by 7%.

Exercise 7.5

Rather than expecting a reduction of 7% in the previous exercise, you are told that the treatment is expected to reduce the number of crashes by 15% with a statistical significant level of 5%. How many years of data would you need to collect in the after period?

First, use Eq. (7.64) again and, second, input the known variables into the equation.

continued

Exercise 7.5 (*cont'd*)

$$\frac{100 - 0.85 \times 90}{\sqrt{((100 \times 4)/4^2 + (90 \times t_a)/t_a^2)}} = 1.96$$

Solve for t_a

$$\frac{100 - 0.85 \times 90}{\sqrt{(400/16 + 90/t_a)}} = 1.96$$

$$\frac{23.5}{\sqrt{25 + 90/t_a}} = 1.96$$

$$\frac{552.25}{25 + 90/t_a} = 3.84$$

$$552.25 = 96 + 345.6/t_a$$

$$456.25 = 345.6/t_a$$

$$t_a = 345.6/456.25 = 0.76 \text{ year}$$

The number of years for the after period is 0.76 year or 9.1 months. In this case, it would probably be recommended to use 10 months, which would cut the collect the data collection efforts by more than one half.

7.8.3 Sample size based on the variance and ratio r_d (before period)

A simpler method for estimating the sample size was proposed by Hauer (1997). With this method, the sample size required for the before period is estimated using the index θ, the significance level sought, and the duration ratio between the after and before time periods r_d, already explained earlier. The method is applicable when the before–after study applies to a treatment group (naïve) and when a comparison group is used.

When the naïve method is used, the following equation can be used (Hauer, 1997):

$$\sum \kappa_i = \frac{\theta/r_d + \theta^2}{\sigma^2(\theta)} \tag{7.66}$$

where $\sum \kappa_i$ is the number of crashes at the treatment sites in the before period; θ is the index; $\sigma^2(\theta)$ is the variance of the index; and, r_d is the ratio of the time difference. Eq. (7.66) shows that a smaller variance or standard

deviation will require more crashes to be collected in the before period, as expected. This equation also shows that the sample size can be reduced for the before period, while still maintaining the same index and variance, if we increase the length of the after time period, which increases r_d.

If a comparison group is included in the analysis, then Eq. (7.67) can be used. This equation add the uncertainty associated with data collected via the comparison group and the variance of odds ratio $Var(\omega)$, as described in Eq. (7.18). This means that more crashes need to be collected at treated sites.

$$\sigma^2\{\theta\} = \frac{\theta/r_d + \theta^2}{\sum \kappa_i} + \theta^2\left[\frac{1/r_d + 1}{\sum \mu_j} + \frac{Var(\omega)}{\omega^2}\right] \tag{7.67}$$

where $\sum \mu_j$ is the number of crashes in the comparison group (before time period); ω is the odds ratio between the treatment and comparison groups; $Var(\omega)$ is the variance of the odds ratio; and, all the other variables are the same as for Eq. (7.66). For this equation, $\sum \mu_j$ needs to be known before hand. Usually, ω is or should be very close to 1 and $Var(\omega)$ is often equal to 0.001 (as described earlier).

Once the sample size is estimated, then the number of sites needs to be determined, given the average crash rates observed at the treated sites.

Exercise 7.6

Estimate the sample size if the treatment is expected to reduce the number of crashes by 15% with an uncertainty of 5% and the same duration for the after and before time periods (4 years each).

First, use Eq. (7.66)

$$\sum \kappa_i = \frac{\theta/r_d + \theta^2}{\sigma^2\left(\widehat{\theta}\right)}$$

Then, include the relevant information into the equation.

$$\sum \kappa_i = \frac{0.85/1 + 0.85^2}{0.05^2}$$

$$\sum \kappa_i \approx 630$$

A total of 630 crashes are needed in the before period.

continued

Exercise 7.6 (*cont'd*)

Let us assume that, on average, 12 crashes per year occur on treated sites and we have 15 sites in the treatment group. The total number of crashes observed is as follows:

$$12 \times 15 \times 4 = 720$$

As we have access to 720 crashes in the before period, we have enough crashes to conduct the analysis.

Exercise 7.7

Using the data from Exercise 7.5, recalculate the sample size if 3000 crashes could be collected at sites that are part of the comparison group. Assume $\omega = 1$ and $Var(\omega) = 0.001$.

First, use Eq. (7.67).

$$\sigma^2\{\theta\} = \frac{\theta/r_d + \theta^2}{\sum \kappa_i} + \theta^2\left[\frac{1/r_d + 1}{\sum \mu_j} + \frac{Var(\omega)}{\omega^2}\right]$$

Then, include the information in the right-hand side of the equation.

$$\theta^2\left[\frac{1/r_d + 1}{\sum \mu_j} + \frac{Var(\omega)}{\omega^2}\right]$$

$$0.85^2\left[\frac{1/1 + 1}{3000} + \frac{0.001}{1}\right]$$

$$0.7225 \times [0.00067 + 0.001]$$

$$0.7225 \times 0.00167$$

$$0.0012$$

Solve for the rest of the equation:

$$0.05^2 = \frac{0.85/1 + 0.85^2}{\sum \kappa_j} + 0.0012$$

$$0.0025 = \frac{1.57}{\sum \kappa_i} + 0.0012$$

$$\sum \kappa_i = \frac{1.57}{0.0013} \approx 1210$$

Exercise 7.7 (*cont'd*)

If a comparison group is used, the number of crashes needed at treated sites almost double (from 630 to 1210 crashes). If no other treated sites can be used, then the comparison group may not be an option (a common occurrence that has been experienced by the authors of this textbook), as the effect of the treatment may not be statistically significant.

For the EB method, the sample size will be governed by two factors: (1) the number of sites in the treatment group, and (2) the sample size needed to estimate a reliable NB regression model. For the former, Eq. 7.66 needs to be used, while for the latter, the values documented in Table 6.3 in Chapter 6 - *Cross-sectional and Panel Studies in Safety*, which describes the sample size needed as a function of the sample mean of the crashes for the sites found in the comparison group, should be used.

For the FB method, the minimum sample size is described in Table 6.4 in Chapter 6—*Cross-Sectional and Panel Studies in Safety*. The minimum sample size was initially developed for Poisson-lognormal models but is also applicable for other mixed-Poisson models.

References

AASHTO, 2010. Highway Safety Manual, first ed. American Association of State Highway and Transportation Officials, Washington, DC.

Abbess, C., Jarrett, D., Wright, C.C., 1981. Accidents at blackspots: estimating the effectiveness of remedial treatment, with special reference to the "regression-to-the-mean" effect. Traffic Eng. Contr. 22, 535–542.

Allison, P.D., 2012. Survival Analysis Using SAS: A Practical Guide. SAS Institute, Cary, NC.

Ash, J.E., Zou, Y., Park, B.-J., Lord, D., Wang, Y., 2017. Application of Finite Mixture of Negative Binomial Regression Models in Estimating Empirical Bayes Estimates. J. Appl. Stat. Vol. 45 (9), 1652–1669.

Aul, N., Davis, G., 2006. Use of propensity score matching method and hybrid Bayesian method to estimate crash modification factors of signal installation. Transp. Res. Rec. 1950, 17–23.

Austin, P.C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavior. Res. 46 (3), 399–424.

Biondi-Zoccai, G., Romagnoli, E., Agostoni, P., Capodanno, D., Castagno, D., D'Ascenzo, F., Sangiorgi, G., Grazia Modena, M., 2011. Are propensity scores really superior to standard multivariable analysis? Contemp. Clin. Trials 32 (5), 731–740.

Copas, J., 1997. Using regression models for prediction: shrinkage and regression to the mean. Stat. Methods Med. Res. 6, 167–183.

Davis, G., 2000. Accident reduction factors and causal inference in traffic safety studies: a review. Accid. Anal. Prev. 32, 95–109.

FHWA, 2009. Manual on Uniform Traffic Control Devices for Streets and Highways. Federal Highway Administration, Washington, D.C.
Hauer, E., 1980a. Bias-by-selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment. Accid. Anal. Prev. 12, 113–117.
Hauer, E., 1980b. Selection for treatment as a source of bias in before-and-after studies. Traffic Eng. Control 21, 419–421.
Hauer, E., 1986. On the estimation of the expected number of accidents. Accid. Anal. Prev. 18, l–12.
Hauer, E., 1997. Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Pergamon Press, Elsevier Science, Ltd., Oxford, United Kingdom.
Hauer, E., 2008. How many accidents are needed to show a difference? Accid. Anal. Prev. 40 (4), 1634–1635.
Hauer, E., Persaud, B., 1983. Common bias in before-and-after accident comparisons and its elimination. Transport. Res. Rec. 905, 164–174.
Hauer, E., Persaud, B., 1984. Problem of identifying hazardous locations using accident data. Transport. Res. Rec. 975, 36–43.
Ibrahim, J.G., Chen, M.-H., Sinha, D., 2001. Bayesian Survival Analysis. Springer, New York, NY.
Kelsey, J.L., Thompson, W.G., Evans, A.S., 1986. Methods in Observational Epidemiology. Oxford University Press, Inc., New York, USA.
Kuo, P.-F., Lord, D., 2013. Accounting for site-selection bias in before-after studies for continuous distributions: characteristics and application using speed data. Transport. Res. Pol. Pract. 49, 256–269.
Kuo, P.-F., Lord, D., 2017. Estimating the safety impacts in before-after studies using the naïve adjustment method. Transportmetrica: Transport. Sci. 13 (10), 915–931.
Li, L., Donnell, E.T., 2020. Incorporating Bayesian methods into the propensity score matching framework: a no-treatment effect safety analysis. Accid. Anal. Prev. 145, 105691.
Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical modelsin road safety countermeasure effectiveness studies in Iowa. Accid. Anal. Prev. 40, 1531–1542.
Li, H., Graham, D.J., Ding, H., Ren, G., 2019. Comparison of empirical Bayes and propensity score methods for road safety evaluation: a simulation study. Accid. Anal. Prev. 129, 148–155.
Lim, S., Marcus, S.M., Singh, T.P., Harris, T.G., Seligson, A.L., 2014. Bias due to sample selection in propensity score matching for a supportive housing program evaluation in New York City. PloS One 9 (10), e109112.
Lord, D., 2006. Modeling motor vehicle crashes using poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prev. 38 (4), 751–766.
Lord, D., Kuo, P.F., 2012. Examining the effects of site selection criteria for evaluating the effectiveness of traffic safety improvement countermeasures. Accid. Anal. Prev. 47, 52–63.
Miranda-Moreno, L., Fu, L., Ukkusuri, S., Lord, D., 2009. How to incorporate accident severity and vehicle occupancy into the hotspot identification process? Transport. Res. Rec. 2102, 53–60.
Park, E.S., Park, J., Lomax, T.J., 2010. A fully Bayesian multivariate approach to before–after safety evaluation. Accid. Anal. Prev. 42, 1118–1127.
Pawlovich, M.D., Wen, L., Carriquiry, A., Welch, T., 2006. Iowa's experience with road diet measures: use of Bayesian approach to assess impacts on crash frequencies and crash rates. Transport. Res. Rec. 1953, 163–171.
Persaud, B., Lyon, C., 2007. Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions. Accid. Anal. Prev. 39, 546–555.

Persaud, B.N., Retting, R., Garder, P., Lord, D., 2001. Observational before-after study of U.S. roundabout conversions using the empirical Bayes method. Transp. Res. Rec. 1751, 1–8.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Sasidharan, L., Donnell, E.T., 2013. Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: Exploratory analysis using intersection lighting data. Accid. Anal. Prev. 539–553.
Shirazi, M., Lord, D., Geedipally, S.R., 2016. Sample-Size Guidelines for Recalibrating Crash Prediction Models: Recommendations for the Highway Safety Manual, Vol. 93, pp. 160–168. Accid. Anal. Prev.
Stingler, S., 1997. Regression towards the mean, historically considered. Stat. Methods Med. Res. 6, 103–114.
Winkelmayer, W.C., Kurth, T., 2004. Propensity scores: help or hype? Nephrol. Dial. Transplant. 19 (7), 1671–1673.
Wood, J.S., Donnell, E.T., 2017. Causal inference framework for generalizable safety effect estimates. Accid. Anal. Prev. 104, 74–87.
Wood, J.S., Donnell, E.T., Porter, R.J., 2015. Comparison of safety effect estimates obtained from empirical Bayes before–after study, propensity scores-potential outcomes framework, and regression model with cross-sectional data. Accid. Anal. Prev. 75, 144–154.
Xie, K., Ozbay, K., Yang, H., Yang, D., 2019. A new methodology for before–after safety assessment using survival analysis and longitudinal data. Risk Anal. 39 (6), 1342–1357.
Zou, Y., Lord, D., Zhang, Y., 2013. Comparison of Sichel and negative binomial models in estimating empirical Bayes estimates. Transport. Res. Rec. 2392, 11–21.