Study Design in Highway Safety

Originally Prepared
by
Dr. Byung-Jung Park & D. Lord

Partially covered in Chapters 6 and 7 of textbook

Fall 2021

Types of Studies

Descriptive

- Case reports
- Case series
- Qualitative studies and surveys

Analytic

Experimental

- (Randomized) Parallel group
- (Randomized) Cross-over

Observational

- Cohort study
- Case-control study
- Cross-sectional study
- Before-After Study

Differences between cohort and case-control studies:
"Observational Studies: Cohort and Case-Control Studies"
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2998589/

Experimental vs Observational

Experimental Study

- The study where the conditions are under the direct control of the researcher. Usually this study involves giving a group of units/individuals a treatment that would not have occurred naturally. Such studies are often used to test the effects of a treatment in people and usually involve comparison with a group that do not get the treatment (laboratory experiments, field experiments).

Observational Study

- This study is one where the researchers have no control over exposures and instead observe what happens to groups of entities (intersections, roadway segments, etc.).
- Assignment of treatments to the entities is not made by the researchers.
- Four types described earlier.

Goal of Study Design

Goal of Study Design

- To obtain valid data so that the research objective(s) can be fulfilled.
  - Research objective(s): to distinguish between the true effect of a treatment and the effect of many other factors
- In highway safety: to use as few units (entities or accidents) as possible for a given degree of precision.

In Observational Study

- Data for observational study are either collected by the researcher for the purpose of the study, or have already been collected for another purpose but is used by the researcher to examine a new research question (May suffer from selection bias).

Challenges

Determining the Sample Size

When planning a study to compare the effectiveness of a particular treatment, it is important to ensure that the sample is of an adequate size to ensure that there is a reasonable chance a clear answer can be produced by the study.

Evaluating the Validity of the Study

- Internal validity: refers to confidence that the findings from a given study are attributable to the treatment alone.
- External validity: refers to the generalizability of findings to other populations, settings, or occasions.
- There are various threats that can invalidate the study design (will be discussed later).

Sample Size

Rule of Thumb

- Make use of the basic principle of inferential statistics that of the normal distribution

$$P\left(\hat{\theta}-1\cdot\sigma(\hat{\theta})\le\theta\le\hat{\theta}+1\cdot\sigma(\hat{\theta})\right)\approx 65\%$$

$$P\left(\hat{\theta}-2\cdot\sigma(\hat{\theta})\le\theta\le\hat{\theta}+2\cdot\sigma(\hat{\theta})\right)\approx 95\%$$

$$P\left(\hat{\theta}-3\cdot\sigma(\hat{\theta})\le\theta\le\hat{\theta}+3\cdot\sigma(\hat{\theta})\right)\approx 99.9\%$$

Four Factors that Need to be Considered

- Variance of the variable being studied
- Size of the effect of interest
- Level of significance (related to type I error)
- Power of a test (related to type II error)

Variance

- Its square root is either standard deviation or standard error
- Standard Deviation: the measure of how variable individual observations are in a sample
- Standard Error: the measure of how variable the mean or proportion is from one sample to another

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

Size of Effect

- The expected size of an effect should be assumed
- This is usually based on the results of previous or pilot studies
- Example
  - A treatment is thought to reduce the expected number of accidents by 10% (i.e., $\theta = 0.9$)

Significance Level

- The significance level tells us how likely it is that an observed difference is due to chance when the true difference is 0.

  H_0: $\theta_1 = \theta_2$ (no difference)

  H_A: $\theta_1 - \theta_2 > 0$

	Do not reject H_0	**Reject H_0**
H_0 is True	Correct Decision 1-α: Confidence level	Type I error α: Significance level
H_0 is False	Type II error β	Correct Decision 1-β: Power of a test

- Sample size can be determined by considering the significance level only.
- However, in order to detect the specific effect of a treatment, the sample size can be determined by considering both significance level and power.

Example 1:

On a certain kind of road on which there are 1.5 reported accidents/km-year an intervention is contemplated. The question is how many kilometres of road are needed so that one can be 95% confident that in a before-after study a 10% reduction in expected accident frequency is detected if 3 years of 'before' and 1 year of 'after' data will be used.

Solution:

Let, x_1, x_2 = accident counts for c_1 and c_2 years on n kilometres of road

Subscript 1 and 2 represents 'before' and 'after' period

Then, $x_1=1.5*3*n=4.5n$

$x_2=(1.5)*(0.9)*1*n=1.35n$

$$\frac{(x_1/nc_1)-(x_2/nc_2)}{\sqrt{x_1/(nc_1)^2+x_2/(nc_2)^2}}=\frac{(1.5)-(1.35)}{\sqrt{4.5/9n+1.35/n}}\approx 2.0$$

This yields n=330 km.

Therefore, x_1=495 acc/year and x_2=446 acc/year are required.

Source:

Hauer, E. (2008) How many accidents are needed to show a difference? Accid Anal Prev 40(4): 1634-5.

Example 2:

Conversely, you might want to know with what confidence can one say that there was a safety improvement from 'before' to 'after' if the count of accidents during a 3-year before period was 61 and during a 2-year after period 29.

Solution:

$$\frac{(x_1/c_1)-(x_2/c_2)}{\sqrt{x_1/(c_1)^2+x_2/(c_2)^2}}=\frac{(61/3)-(29/2)}{\sqrt{61/9+29/4}}\approx 1.55 \lim_{x\to\infty}$$

Since 1.55 < 2, the hypothesis that $\mu_1 = \mu_2$ would not be rejected in favor of the hypothesis that $\mu_2 < \mu_1$ at a level of significance at 5%.

Source:

Hauer, E. (2008) How many accidents are needed to show a difference? Accid Anal Prev 40(4): 1634-5.

Power of a Test

- Power is the probability that it will correctly lead to the rejection of a false null hypothesis.
- We can think of power as the probability of detecting a true effect.
- Two different aspects of power analysis. One is to calculate the necessary sample size for a specified power. The other aspect is to calculate the power for given a specific sample size.
- Generally, a test with a power greater than 0.8 (or $\beta <= 0.2$) is considered statistically powerful.

LD1

$$H_0: \mu_1 - \mu_2 = d \quad vs. \quad H_A: \mu_1 - \mu_2 = d^*$$

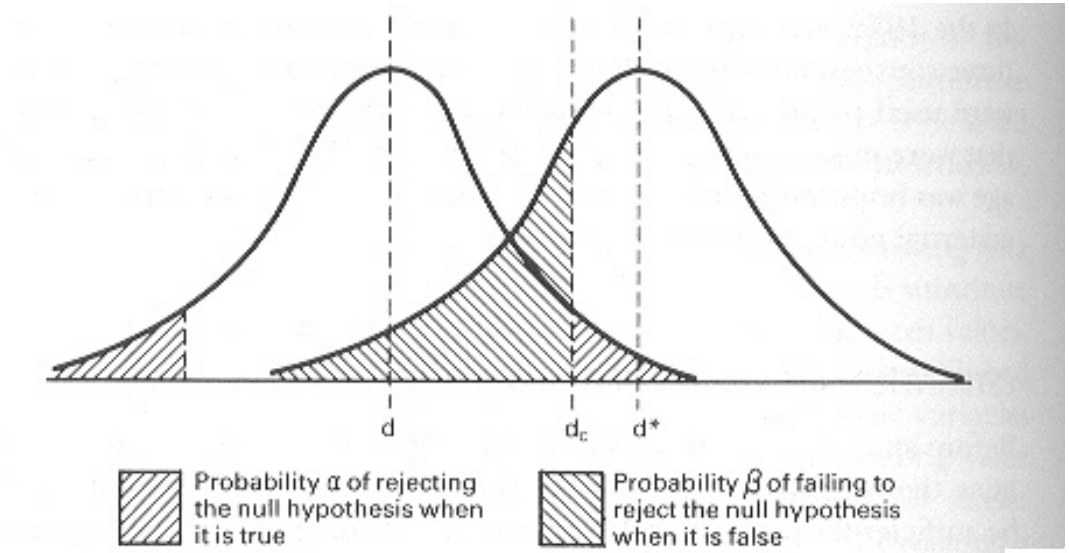


Figure 10–3. Sampling distribution of a measure of association when the null hypothesis i true and when it is false. The symbol d denotes the null value of the measure of associatio d^* denotes the nonnull value of the measure, and d_c denotes the value of the measure th is just significant at the significance level α.

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

In textbook

$$d_c = d + Z_{\alpha/2} s.e.(d) = d^* - Z_\beta s.e.(d^*)$$

$$Z_\beta = \frac{d^* - d - Z_{\alpha/2} s.e.(d)}{s.e.(d^*)} = \frac{d^* - d}{s.e.(d^*)} - Z_{\alpha/2}$$

Slide 12

LD1 Lord, Dominique, 11/29/2021

Reconsider Example 1:

What is the power of the test to detect the difference in expected accident frequency between before and after periods, $\mu_1-\mu_2=0.15$? Use significance level $\alpha=0.05$.

Solution:

$$Z_\beta = \frac{d^* - d}{s.e.(d^*)} - Z_{\alpha/2} = \frac{0.15-0}{\sqrt{\frac{4.5}{(9)(330)} + \frac{1.35}{(1)(330)}}} - 1.96$$

$$= 0.043$$

Find the power (1-β) which corresponds to $Z_\beta=0.043$.

Statistical Power ≈ 51.5%

How can we increase the power?

Continuing…

How many kilometres of road are needed so that one can be 95% confident in a 10% reduction in expected accident frequency, while ensuring the statistical power of 80% ?

Solution:

$$\frac{0.15-0}{\sqrt{4.5/9n+1.35/n}} = Z_{\alpha/2} + Z_{\beta} = 2.80$$

This yields n=644.6 km.

Therefore, x_1=967 acc/year and

x_2=871 acc/year are required.

Almost doubled up!!

Table 10–16. Values of $(Z_{\alpha/2} + Z_{\beta})^2$ for frequently used combinations of significance level and power

Significance level (α)	Power ($1-\beta$)	$(Z_{\alpha/2} + Z_{\beta})^2$
0.01	0.80	11.679
	0.90	14.879
	0.95	17.814
	0.99	24.031
0.05	0.80	7.849
	0.90	10.507
	0.95	12.995
	0.99	18.372
0.10	0.80	6.183
	0.90	8.564
	0.95	10.822
	0.99	15.770

Study Design for Before-After Studies

Naïve Before-After Study

Using a Comparison Group

Empirical Bayes Approach

Naïve Before-After Study

Two decisions that need to be made

- The number of entities (or accidents) for the treatment group
- The duration of the 'before' and 'after' periods

Precision = Standard error of the estimate, $\sigma(\hat{\theta})$

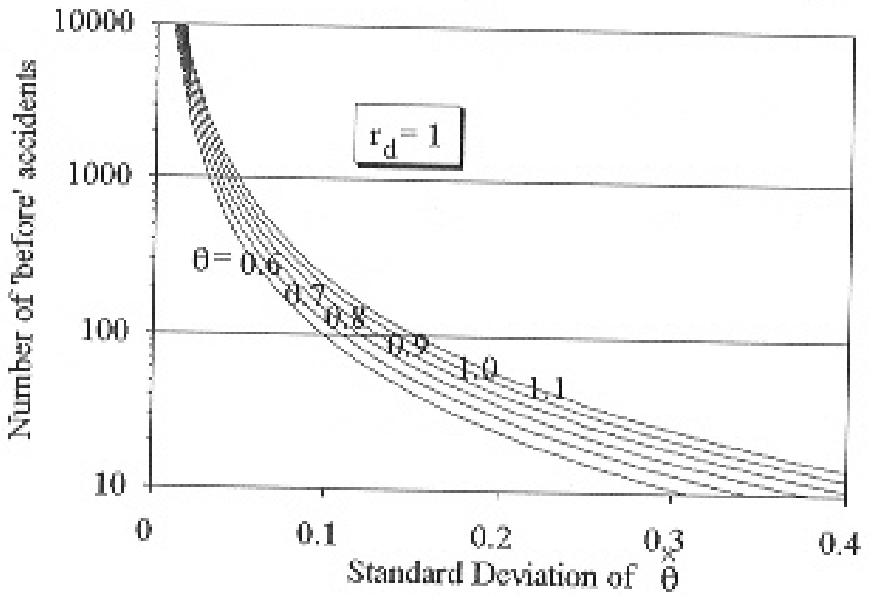
$$\sum \kappa(j) = \frac{\theta / r_d + \theta^2}{\sigma^2(\hat{\theta})} \approx \frac{2}{\sigma^2(\hat{\theta})}$$

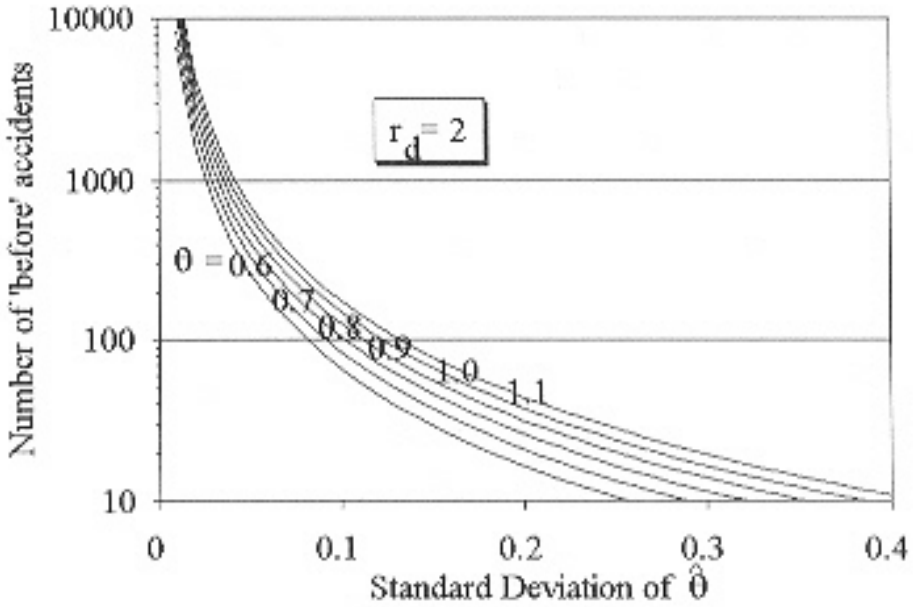
$$P\left(|\hat{\theta} - \theta| \le 1 \cdot \sigma(\hat{\theta}) \right) = 65\% \qquad P\left(|\hat{\theta} - \theta| \le 2 \cdot \sigma(\hat{\theta}) \right) = 95\%$$

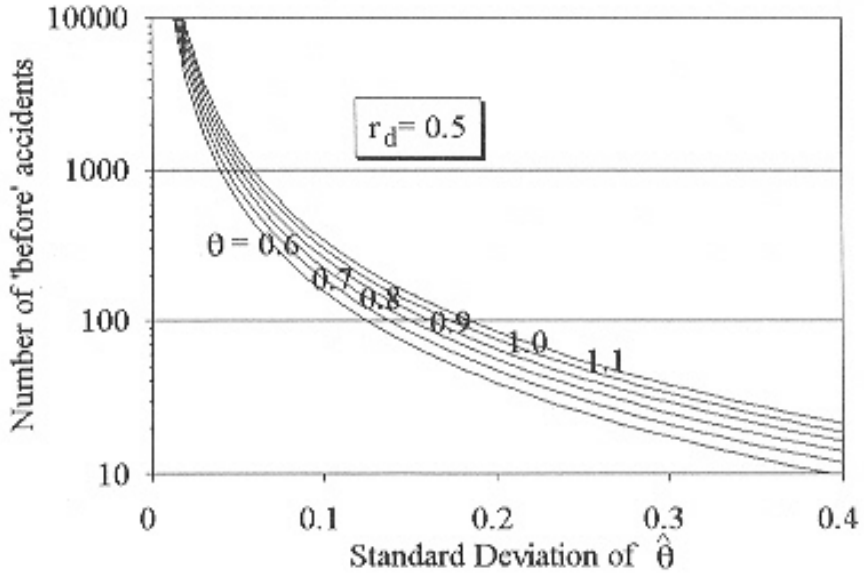
When $\sigma(\hat{\theta}) = 0.1$, we need 200 'before' accidents

$\sigma(\hat{\theta}) = 0.01$, we need 20,000 'before' accidents

$$\sum \kappa(j) = fn\left\{ \theta, r_d, \sigma(\hat{\theta}) \right\}$$







Example: A treatment is thought to reduce the expected number of crashes by 10% (i.e., $\theta = 0.9$). If the before and after period are one year in duration, what is the number of crashes need for the before period for $\sigma(\hat{\theta}) = 0.05$?

$$\sum \kappa(j) = \frac{0.9/1 + 0.9^2}{0.05^2} \approx 700 \text{ crashes}$$

What if the system can provide only 175 accidents per year?
How can we get the same statistical precision $\sigma(\hat{\theta}) = 0.05$?

- Option 1: Increase the 'before' and 'after' periods to 4 years
- Option 2: Increase the 'before' period to 3 years, and the 'after' period to 5.4 years

Are those good options? (see the five naïve assumptions in page 74)

Using a Comparison Group

The sample size needed when the study includes a control group, is governed by the terms $\sigma^2\{\hat{\theta}\}$ or $Var\{\theta\}$ and $Var\{\omega\}$

$$\sigma^2\{\hat{\theta}\} = \frac{\theta / r_d + \theta^2}{\sum \kappa(j)} + \theta^2 \left[\frac{1/r_d + 1}{\sum \mu(j)} + \frac{Var(\omega)}{\omega^2} \right]$$

Number of crashes in treatment group

Number of crashes in control group

Variance of odd ratios

odd ratios (usually close to 1)

This is estimated from the control and treatment groups

Example: Taking the same example as before with $\sigma\{\hat{\theta}\} = 0.05$, now assume the control group contains 5,000 crashes for the before period with $Var(\omega) = 0.001$ and $\omega = 1.0$

The comparison group contributes to the overall variance

$$\theta^2 \left[\frac{1/r_d + 1^2}{\sum \mu(j)} + \frac{Var(\omega)}{\omega^2} \right] = 0.9^2 \left[\frac{2}{5,000} + 0.001 \right] = 0.0011$$

$$\sigma^2\{\hat{\theta}\} = 0.0025 = \frac{\theta/r_d + \theta^2}{\sum \kappa(j)} + 0.0011 = 0.0014$$

$$\frac{\theta/r_d + \theta^2}{\sum \kappa(j)} = 0.0014 \qquad \sum \kappa(j) = \frac{0.9/1 + 0.9^2}{0.0014} = 1,222 \text{ crashes}$$

Empirical Bayes Approach

$$\mu_{EB} = w \times \mu + (1 - w) \times y$$

μ_{EB} = Estimate of the expected number of crashes for an entity of interest

μ = Expected number of crashes based on expected on similar entities

y = number of crashes on the entity of interest

w = Weight factor $= \dfrac{1}{1 + \mu / \phi}$

- The sample size issue arises when μ is estimated from a statistical model (a negative binomial model)
- Larger sample size reduces the bias in the dispersion parameter estimate (see next two slides)
- Given the characteristics of crash data, i.e. Low mean and overdispersion, models should be developed with at least 100 observations. Ideally, more than 1,000 observations should be used.

Sample Size

TABLE 6.4 Recommended sample size (Lord, 2006).

Population sample mean	Minimum sample size
5.00	200
4.00	250
3.00	335
2.00	500
1.00	1000
0.75	1335
0.50	2000
0.25	4000

NB models estimated using the MLE

Sample Size

TABLE 6.5 Recommended minimum sample size for Bayesian Poisson-lognormal models (Miranda-Moreno et al., 2008).

Population sample mean	Minimum sample size
≥2.00	20
1.00	100
0.75	500
0.50	1000
0.25	3000

NB/PLN models estimated using the Bayesian method

(Note: if using the FB, there is no need to use the EB)

References on the effect of small sample size on the negative binomial dispersion parameter estimate

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*. Accident Analysis & Prevention, Vol. 38, No. 4, pp. 751-766.

Lord, D., and L.F. Miranda-Moreno (2008) *Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective*. Safety Science, Vol. 46, No. 5, pp. 751-770.

Park, B.-J., and D. Lord (2008) *Adjustment for the Maximum Likelihood Estimate of the Negative Binomial Dispersion Parameter*. Transportation Research Record 2061, pp. 9-19.

Threats to Validity of Study Design

- Observational studies are evaluated in terms of both internal and external validity.
- All possible threats to validity cannot be controlled in any one study
- Cook and Campbell (1979) identified the threats to the validity of designs.

Cook TD & Campbell DT (1979) Quasi-experimentation: design and analysis issues in field settings. Chicago, Rand McNally

Threats to internal validity

Threats	Features
a) History	Event external to treatment which may affect dependent variable.
b) Maturation	Biological and psychological changes in subjects which will affect their responses.
c) Testing	Effects of pre-test may alter responses on post-test regardless of treatment.
d) Instrumentation	Changes in instrumentation, raters or observers.
e) Statistical regression	Extreme scores tend to move to middle on post-testing regardless of treatment.
f) Selection	Differences in subjects prior to treatment.
g) Mortality	Differential loss of subjects during study.
h) Interaction of selection with maturation, history and testing	Other characteristic of subjects mistaken for treatment effect on post-testing, differential effects in selection factors.
i) Ambiguity about direction of causality	In studies conducted at one point in time, problem inferring direction of causality.
j) Diffusion/imitation of treatment	Treatment group share the conditions of their treatment with each other.
k) Compensatory equalization	It is decided that everyone in experimental or comparison group receive the treatment that provides desirable goods and services.
l) Demoralization of Respondents	Members of the group not receiving the treatment perceive they are inferior and give up.

Note: Threats somewhat related to our safety study are highlighted in red.

Threats to external validity

Threats	Features
a) Interaction of selection and treatment	Ability to generalize the treatment to persons beyond the group studies
b) Interaction of setting and treatment	Ability to generalize to other settings beyond the one studies
c) Interaction of history and treatment	Ability to generalize the treatment to other time beyond the one studies

Note: Threats somewhat related to our safety study are highlighted in red.