Statistical Issues in Road Safety:
Evaluation of Safety Measures

Shrikant I. Bangdiwala, PhD
McMaster University
Hamilton, Canada

Learning objectives

- Understand the use of different study designs available for our research questions of diagnosis or of evaluation
- Understand the methods of statistical inference to ‘See the signal amidst the noise’ and to understand information in the presence of uncertainty from variability and chance
- Understand the principles behind statistical models of relationships that attempt to ‘make sense out of reality’
- Understand how to integrate study designs, statistical models, and the principles of statistical inference for the evaluation of safety measures → to decide if something works or something does not work
Exercise

- Research Question:
  Do lower speeds lead to safer roads?

- How do we answer this question?
  - What type of study?
  - How do we define 'lower'? How do we define 'safer'?
  - Who or what do we study? How many?
  - Who or what do we compare results to? How many?
  - What data do we collect? How do we measure it? When do we measure? For how long do we measure?
  - What is a meaningful relationship?
  - How can we know if what we observe could have been due to chance?

RQ: Do lower speeds lead to safer roads?

- What do you need to do first?
- Must recognize the ‘sources of variability’

Type of road – i.e. 2-way, straight sector, intersection
Traffic volume?
Presence of median?
What other factors? → consider Haddon Matrix
RQ: Do lower speeds lead to safer roads?

- What do you need to do next?
- Choose an intervention – ‘evidence-based’
  - How much lower?
  - Other structural aspects may have a confounding effect!
  - How do you know what works?

Evidence that lower speeds lead to fewer road transport crashes?

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>log(RR)</th>
<th>Rate Ratio</th>
<th>Weight</th>
<th>Rate Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(SE)</td>
<td>95% CI</td>
<td></td>
<td>95% CI</td>
</tr>
<tr>
<td>Copenhagen 1990</td>
<td>-0.296</td>
<td>0.75 [0.57, 0.99]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gdansk 1991</td>
<td>-0.618</td>
<td>0.54 [0.32, 0.91]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gothenburg 1991</td>
<td>-0.601</td>
<td>0.53 [0.32, 0.88]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gothenburg 1993</td>
<td>-0.151</td>
<td>0.86 [0.59, 1.24]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gothenburg 1995</td>
<td>-0.229</td>
<td>0.80 [0.50, 1.29]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gothenburg 1997</td>
<td>-0.357</td>
<td>0.70 [0.51, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stockholm 1992</td>
<td>-0.373</td>
<td>0.70 [0.51, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>London 2007</td>
<td>-0.502</td>
<td>0.60 [0.40, 0.88]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malm 1993</td>
<td>-2.096</td>
<td>0.12 [0.02, 0.69]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osaka 1998</td>
<td>-0.143</td>
<td>0.86 [0.54, 1.39]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sydney-Camdenbury 1990</td>
<td>-0.182</td>
<td>0.83 [0.55, 1.25]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sydney-Wilcannia 1990</td>
<td>-0.357</td>
<td>0.70 [0.51, 0.96]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI): 100.0% [0.76, 1.05]

Heterogeneity: Q = 64, df = 11, p = 0.91
Chi-square = 13.3, df = 11, p = 0.91
Test for overall effect Z = 1.38 (p = 0.17)
Evidence that lower speeds lead to fewer road transport injuries?

Next step → choose a study design
Study designs

Descriptive
- Case report
- Case series
- Descriptive epidemiological study
- Ethnographic study
- Other

Analytic
- Experimental
  - Randomized controlled trial
  - Non-randomized (Quasi-experimental)
  - Before-after intervention
  - Comparison study

Observational
- Case-control study
- Case-crossover
- Cross-sectional study
- Longitudinal designs
  - Cohort study
  - Other

Before-after 'natural experiment'
- Interrupted time series
- Other
- Retrospective
- Prospective


Observational studies – in a given time period

<table>
<thead>
<tr>
<th>Type of study</th>
<th>What do they consist of</th>
<th>What information they provide</th>
<th>What statistics they provide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case series</td>
<td>In-depth examination of a few number of cases</td>
<td>Presence or absence of multiple factors in cases</td>
<td>Descriptions of distribution of risk factors</td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td>A sample from a population at a given moment in time</td>
<td>If sample is population based, estimates of the prevalence of the outcome</td>
<td>Correlations for continuous variables; odds ratio of exposure or of outcome for categorical variables</td>
</tr>
</tbody>
</table>
### Observational studies – going back in time

<table>
<thead>
<tr>
<th>Type of study</th>
<th>What do they consist of</th>
<th>What information they provide</th>
<th>What statistics they provide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-control study</td>
<td>Two samples – one of cases, and one of controls</td>
<td>Measures of the strength of the association between ‘caseness’ and multiple potential exposure variables</td>
<td>Odds ratio of exposure</td>
</tr>
<tr>
<td></td>
<td>Retrospective assessment of exposure in both samples</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Case-crossover study
- Within individual or group sample of times prior to the event
- Estimates of the effect of immediate exposures on the occurrence of an acute event or outcome
- Odds ratio of exposure

![Diagram of case-crossover design](image)

### Observational studies – going forward in time

<table>
<thead>
<tr>
<th>Type of study</th>
<th>What do they consist of</th>
<th>What information they provide</th>
<th>What statistics they provide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort study</td>
<td>Two samples – one of exposed individuals, and one of unexposed individuals</td>
<td>If samples are population based, estimates of the incidence of the outcome over time</td>
<td>Relative risk of outcome</td>
</tr>
<tr>
<td></td>
<td>Prospective assessment of occurrence of outcome in both samples</td>
<td>Measures of the risk of multiple outcomes in exposed and unexposed individuals</td>
<td></td>
</tr>
<tr>
<td>Surveillance and registries</td>
<td>All events occurring in a defined population in an extended period of time</td>
<td>Since population based, incidence rates for events*</td>
<td>Descriptions of distributions of events</td>
</tr>
</tbody>
</table>

![Diagram of cohort study](image)
Experimental studies - randomized

<table>
<thead>
<tr>
<th>Type of study</th>
<th>What do they consist of</th>
<th>What information they provide</th>
<th>What statistics they provide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual randomized trial</td>
<td>Individuals are randomly assigned to intervention and control arms; standard protocols are uniformly implemented</td>
<td>Valid estimate of the effect of the exposure (intervention) on primary and secondary outcomes</td>
<td>Relative risks for binary outcomes; effect on means for continuous outcomes</td>
</tr>
<tr>
<td>Group or community randomized controlled trial</td>
<td>Groups of individuals are randomly assigned to intervention and control arms; standard protocols are uniformly implemented</td>
<td>Valid estimate of the effect of the exposure (intervention) on primary and secondary outcomes</td>
<td>Relative risks for binary outcomes; effect on means for continuous outcomes</td>
</tr>
</tbody>
</table>

Novel group randomized designs for injury research

- Cluster cross-over: reversible policy
  - Eligible communities: N
  - Policy A
  - Policy B
  - Randomization:
    - Int. A: Event
    - Int. B: Event
    - No event

- Stepped-wedge: environmental changes
  - Eligible communities: 6
  - At each time period, randomly select next community for Int.

<table>
<thead>
<tr>
<th>T0</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Control</td>
<td>Int.</td>
<td>Int.</td>
<td>Int.</td>
<td>Int.</td>
<td>Int.</td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Int.</td>
<td>Int.</td>
<td>Int.</td>
<td>Int.</td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Int.</td>
<td>Int.</td>
<td>Int.</td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Int.</td>
<td>Int.</td>
</tr>
<tr>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Control</td>
<td>Int.</td>
</tr>
</tbody>
</table>
Experimental studies – non-randomized

<table>
<thead>
<tr>
<th>Type of study</th>
<th>What do they consist of</th>
<th>What information they provide</th>
<th>What statistics they provide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-randomized</td>
<td>Groups or individuals receive different interventions, but are not randomly assigned; they are self-selected to exposure</td>
<td>Biased estimate of the effect of the exposure (intervention) on primary and secondary outcomes</td>
<td>Relative risks for binary outcomes; effect on means for continuous outcomes</td>
</tr>
<tr>
<td>comparison study</td>
<td>A single individual or group acts as its own counterfactual – the outcome after the intervention is compared to the outcome before</td>
<td>Within-person or within-group changes due to the exposure (intervention)</td>
<td>Mean change score for continuous outcomes; McNemar's test for binary outcomes</td>
</tr>
<tr>
<td>RQ: Do lower speeds lead to safer roads?</td>
<td>What do you need to do next? → determine required sample size</td>
<td>How 'fewer' is meaningful?</td>
<td>How to account for variability and the role of chance? → sample size !! Depends on the type of research question:</td>
</tr>
</tbody>
</table>

For estimation

The necessary sample size $n$ when estimating the effect of an intervention in a given group is

- higher if want higher level of confidence
- higher if want higher precision
- lower half-width of confidence interval

Conceptually

$$n \propto \frac{\text{(confidence) (variability)}}{\text{(width)}}$$

For testing

The necessary sample size $n$ when testing the effect on an intervention is

- higher if want lower level of significance
- higher if want higher power
- higher if differences to detect are smaller
- higher if there is more variability

Conceptually

$$n \propto \frac{\text{(power) (variability)}}{\text{(significance) (differences)}}$$
What do you need to do next

- **Do the study** – obtain sample, make observations or measure outcomes
- **Decide analytic methodology (prior to seeing data)**
  - Based on study design
  - Based on types of variables
  - Based on research question/hypothesis!
- **Check data quality**
  - Missingness – missing completely at random (MCAR), missing at random (MAR), missing not at random (MNAR)
  - Errors – out of range, inconsistencies
  - Outliers – unusual values
  - Distributional assumptions
- **Proceed with analyses** – descriptive, inferential

---

Statistical inference

- **Estimation** of a parameter in a sample → subject to sampling error, measurement error
  - Single estimate = ‘point estimate’
  - Interval estimate = ‘confidence interval’
- **Inference** from the sample information to knowledge about a population → based on the design, but subject to sampling error, measurement error, and chance
  - We quantify this chance by assuming a probability model
    - Using indirect proof → formal statistical hypothesis testing
    - Using imprecision estimates → construction of confidence intervals
Confidence intervals

**Limits on precision:**

- Specify a lower bound and an upper bound within which we are highly confident the true (unknown) value of the parameter lies
- The interval is centered around our point estimate
- The width of the confidence interval
  - Increases with our desire for bigger confidence
  - Is larger if there is larger variability (‘noise’ in the system)
  - Is smaller if based on larger sample sizes

---

**Example**

Are 16-year-old drivers with teenager passengers at higher risk of a crash than without other passenger?

```
Relative risk = \frac{T_{16}A_{16}}{T_{16}A_{16}f}.
```

where

- \(T\) = number of crash involvements for the target age driver (e.g. 16-year-old drivers),
- \(A\) = number of crash involvements for adult drivers aged 30–59 (i.e. the base driver group),
- \(f\) = at-fault involvements, and
- \(n_f\) = not-at-fault involvements.

**Table 3. Relative risk of fatal crash involvement by driver age and passenger presence (FARS, 1990–1995)**

<table>
<thead>
<tr>
<th>Driver age</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20–24</th>
<th>25–29</th>
<th>30–54*</th>
<th>65–69</th>
<th>70+</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>3.28</td>
<td>3.67–2.51</td>
<td>2.28</td>
<td>90–106</td>
<td>1.63–1.99</td>
<td>4.45–1.55</td>
<td>1.10</td>
<td>1.06–1.07</td>
<td>2.99</td>
</tr>
<tr>
<td>Driver alone</td>
<td>2.28</td>
<td>1.77</td>
<td>1.50</td>
<td>2.45–1.55</td>
<td>2.54</td>
<td>2.45–1.55</td>
<td>1.00</td>
<td>1.13</td>
<td>2.27</td>
</tr>
<tr>
<td>95% confidence intervals</td>
<td>2.07–2.38</td>
<td>1.85–1.90</td>
<td>1.30–1.72</td>
<td>2.23</td>
<td>2.30–1.38</td>
<td>6.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With passengers</td>
<td>4.72</td>
<td>3.52</td>
<td>3.66</td>
<td>2.54</td>
<td>2.45</td>
<td>2.45</td>
<td>2.64</td>
<td>2.64</td>
<td>1.84</td>
</tr>
<tr>
<td>95% confidence intervals</td>
<td>4.32–5.11</td>
<td>3.68–4.98</td>
<td>3.26–1.93</td>
<td>3.22</td>
<td>3.40–2.45</td>
<td>0.87–0.96</td>
<td>0.84–2.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The 30–59 age group is the reference group for relative risk calculations.

Preusser et al. (1998) The effect of teenage passengers on the fatal crash risk of teenage drivers. Accident Analysis & Prevention
Statistical hypothesis testing

**Indirect proof:**
- Specify a ‘null’ hypothesis (‘status quo’) and an ‘alternative’ hypothesis (our claim)
  - e.g. risk of 16-year old drivers with teenage passengers
    - \( H_0 \): relative risk = 1
    - \( H_A \): relative risk > 1
- Assume the null is correct
- Observe reality (collect data) and see if it conforms with the null being in operation →
  - If probability of observing more \( H_A \) favorable data than what we found is small, we have evidence to reject the \( H_0 \)
  - If probability of observing more \( H_A \) favorable data than what we found is not small, our data are consistent with \( H_0 \)

---

Hypothetical example:
Is this driver drunk \( \rightarrow \) evidence?

**Hypotheses**
- \( H_0 \): driver is not drunk (innocent)
- \( H_A \): driver is drunk (guilty)

**Example of evidence in increasingly incriminating order**
- Non-erratic driving behavior
- Driving behavior was erratic
- Driving behavior led to collision
- Driver cannot write name
- Driver has slurred speech
- Driver smells of alcohol
- Driver cannot do ‘nose test’
- Driver cannot do ‘standing 4 test’
- Driver’s BAC is high
Hypothetical example: Is this driver drunk → evidence?

- Driver's BAC is high
- Driver cannot do 'standing 4 test'
- Driver cannot do 'nose test'
- Driver smells of alcohol
- Driver has slurred speech
- Driver cannot write name
- Driving behavior led to collision
- Driving behavior was erratic
- Non-erratic driving behavior

P-value = probability of an observation that is as or more 'extreme' than the one observed, under the null hypothesis

Summary of statistical inference

- In road safety research, we use statistical methods for inference
  - Estimate a characteristic in the population
  - Compare populations
  - Establish a relationship between variables in a population
- Steps in understanding the observed effects/results:
  1. Does it make sense? → agrees with expected results, direction and magnitude
  2. Is it meaningful clinically? → clinical importance or relevance
  3. Could it be explained by biases? → selection, measurement, recall, analytic, …
  4. Could it be explained by confounding?
  5. Can I rule out the role of chance? → statistical significance
Sir R.A. Fisher linked ‘P-value’ with his concept of ‘significance’

- P-value = an index measuring whether the data are compatible with the null hypothesis

![Graph showing P-values]

- Why? From where? → “Convenience”
- ‘Obviously’ the observed difference is likely to be due to chance → not significant
- ‘Obviously’ the observed difference is unlikely to be due to chance → significant
- ‘Strongly indicated that the null hypothesis fails to account for the whole of the facts’
- ‘Need more information or experimentation’

Statistical inference focusing on interpretation of confidence intervals

- Suppose $\delta$ is a difference that would be considered clinically/contextually important (on a population level or individual level)

- Clinically importantly better, statistically significant
- Uncertain clinical importance, statistically non-significant
- Uncertain clinical importance, statistically significant
- Uncertain clinical importance, statistically non-significant

- The Significance of a statistic is not related to the Importance of the statistic

- Uncertain clinical importance, statistically non-significant
- Uncertain clinical importance, statistically significant
- Clinically importantly worse, statistically significant
Dealing with multiple factors

- **By study design**
  - Experimental designs – control most factors, let few (one) vary
  - Observational designs – control few factors, most do vary

- **By analysis methods**
  - Stratified analysis – study within similar groups – not always possible
  - Regression modeling – adjust for variation of several factors

Design vs. Analysis

- When we control a lot of potential explanatory factors by design, it is not necessary to control much in analysis
  - e.g. randomized controlled trials
- But when we cannot control much by design, as in observational studies, we need to control variability at the analysis
  - Stratified analysis
  - Multivariate regression models
Dealing with complexities

- Multiple factors operate on the systems we study
- We build models – approximations to reality
  - Risk models – probability distributions
  - Sampling models – study designs
  - Regression models – complex systems, relationships

"essentially, all models are wrong, but some are useful"

Modeling relationships

- Interest is in studying the relationship between two variables:
  - Y = dependent variable
  - X = independent variable
- How are they related?
  - Association vs causation
  - Directly or indirectly (mediated, confounded)
  - Unencumbered or moderated (modified, interacted)
  - Linearly or non-linearly
  → Complex system!
Techniques used in statistical modeling

- The world is complex!!
  - Relationships may not be linear: but may be approximately linear in a narrow range of $X$
  - There are multiple variables potentially acting on the relationship: control for them by design or by ‘excluding’ the variable by limiting the population studied
  - There is lots of variability: study more so that the ‘signal’ can be detected despite the ‘noise’

Models

- Multiplicative models – linearized to be fitted
  - Simple
    - Fatality rate = deaths/population
    - Motorization = vehicles/population
    - Traffic risk = deaths/vehicles
      \[ D/V = \frac{D}{P} / \frac{V}{P} \]
  - Complex
    - Exponential, mathematical: e.g. Oppe/Koornstra/Lassarre models
    - Linear models with multiple factors: e.g. statistical regression
    - Non-linear with multiple factors: structural equation models (SEM), classification and regression trees (CART), …
    - Artificial Intelligence based – e.g. neural network, gradient boosting, support vector machine (SVM), …
Concepts of a ‘good’ statistical model

- Concept of ‘model should lie among the observed data points’ → goodness of fit
  - Minimizing ‘residuals’ – specifically the sum of squares of residuals
- Concept of ‘model should enhance understanding of the relationship’ → interpretation must be plausible
  - Structural equation models
  - Mathematical relationships based on engineering, physics

Comprehensive model

  - Developed accident prediction models that take into account the effects of multiple design elements (horizontal, vertical, cross sectional, etc.). The roads are two-lane rural roads without medians.

\[
\hat{y} = \text{EXPO} \times \exp(0.65 + 0.14 \text{STATE} - 0.085 \text{LW} - 0.059 \text{SW} \\
+ 0.067 \text{RHR} + 0.0085 \text{DD}) \\
\times \left[ \sum_i \text{WH} \{i\} \exp(0.045 \text{DEG} \{i\}) \right] \\
\times \left[ \sum_j \text{WP} \{j\} \exp(0.44 \text{V} \{j\}) \right] \\
\times \left[ \sum_k \text{WG} \{k\} \exp(0.11 \text{GR} \{k\}) \right]
\]
Width & Speed

- Speed is a primary consideration when evaluating potential adverse impacts of lane width on safety. On high-speed, rural two-lane highways, an increased risk of cross-centerline head-on or cross-centerline sideswipe crashes is a concern because drivers may have more difficulty staying within the travel lane.

- In a reduced-speed urban environment, the effects of reduced lane width are different. On such facilities, the risk of lane-departure crashes is less. The design objective is often how to best distribute limited cross-sectional width to maximize safety for a wide variety of roadway users. Narrower lane widths may be chosen to manage or reduce speed and shorten crossing distances for pedestrians. Lane widths may be adjusted to incorporate other cross-sectional elements, such as medians for access control, bike lanes, on-street parking, transit stops, and landscaping.

Linear, logistic, Poisson and Cox models

- If outcome is measured on a continuous scale – e.g. AIS score
  \[ Y_{AIS} = \beta_0 + \beta_1(speed) + \beta_2(protection\ gear) + \beta_3(age) + e \]

- If outcome is a proportion between 0-100 – e.g. probability of seat belt use
  \[ \ln(OddsofRisk) = \ln[Pr(Use = 1) / Pr(Use = 0)] = \beta_0 + \beta_1(fine$) + \beta_2(gender) + e \]

- If outcome is a count of events – e.g. number of crashes in a given area over a certain period
  \[ \ln(Y_{avg}) = \beta_0 + \beta_1(traffic\ mix) + \beta_2(road) + e \]

- If outcome is time to the occurrence on an event – e.g. time to first crash of newly licensed males
  \[ Risk = Pr(Y_{crash} < t) = 1 - e^{-\int_0^t h(u)du} \quad \text{where} \quad \ln h(t) = \ln h_0(t) + \beta_0 + \beta_1(training) + e \]
Key points

- Life is complex → Transport systems are complex
- Must understand sources of variability that lead to uncertainty → statistics
- Statistical considerations determine research study designs appropriate to answer specific research questions
  - Types of research questions:
    - Diagnosis → observational studies
    - Evaluation → experimental studies

Key points

- Statistical methods for inference help us
  - ‘See the signal amidst the noise’
  - Understand information in the presence of uncertainty from variability and chance
  - Assuming data is of quality!
- We construct models to ‘make sense out of reality’ (‘what structures gave us these data’)
- All models are simplifications of reality → have limitations stemming from assumptions
  - Statistical models incorporate uncertainty (chance) and variability
- Models to evaluate effectiveness of interventions
  - Need appropriate study design
References


Thanks!