Introduction to road safety: Basic concepts, data and statistical analysis

Luis Miranda-Moreno, PhD
Assistant Professor
Department of Civil Engineering
E-mail: *luis.miranda-moreno@mcgill.ca*

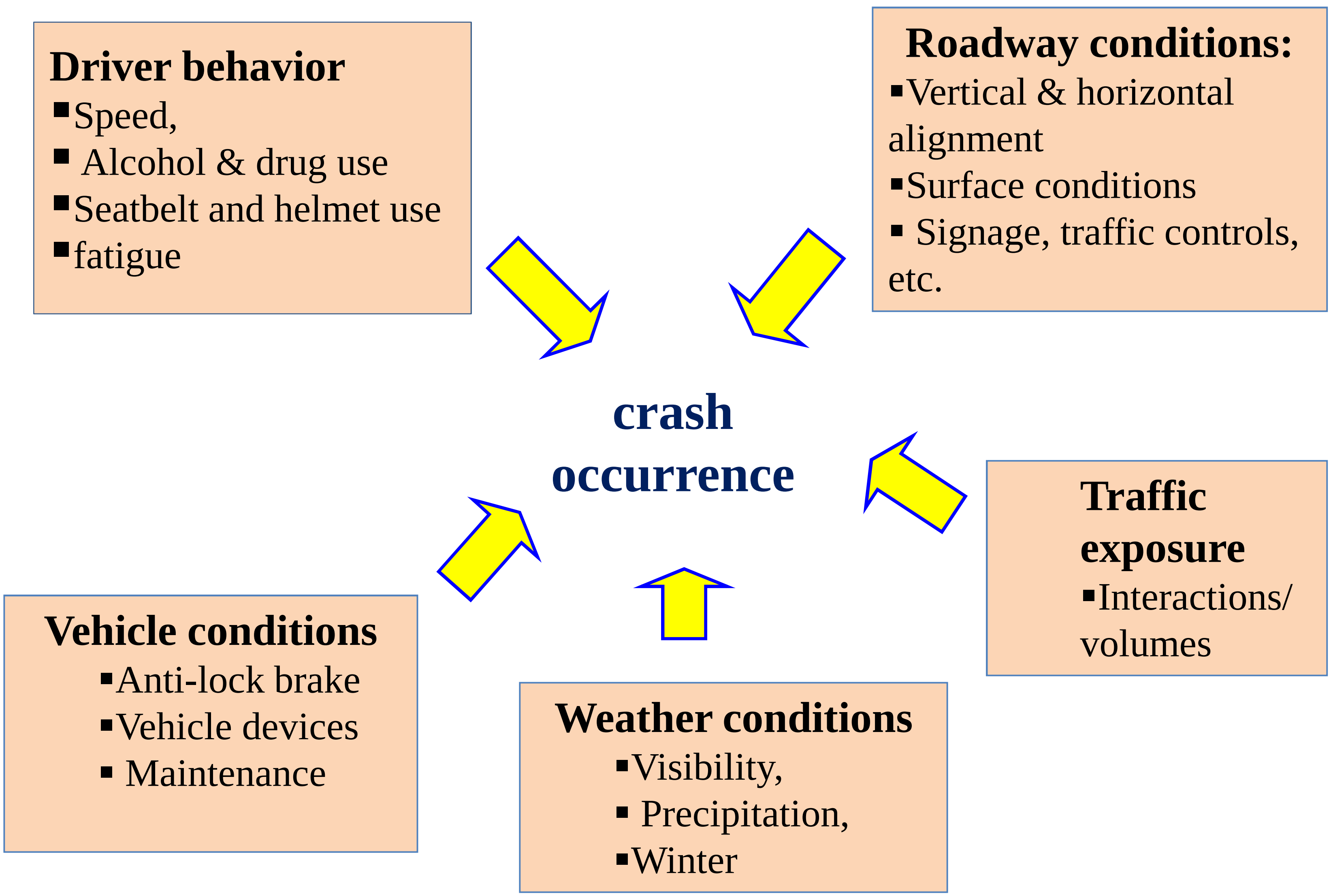
Outline

1. Basic concepts
2. Data preparation and introduction to road safety methods

1. basic definitions

Factors influencing crash occurrence

Road crashes are multi-factorial events

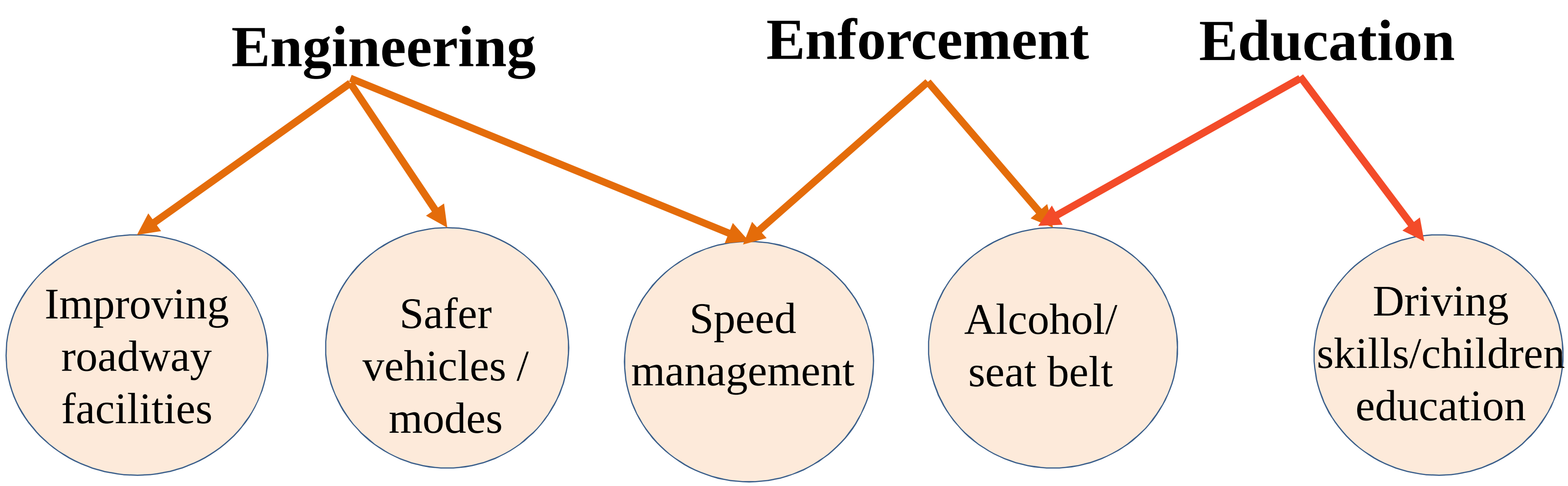


3Es approach

In the "triple E" system, each "E" stands for:

1. Engineering (road engineering and vehicle engineering),
2. Education (training, traffic education) and
3. Enforcement.

3Es approach

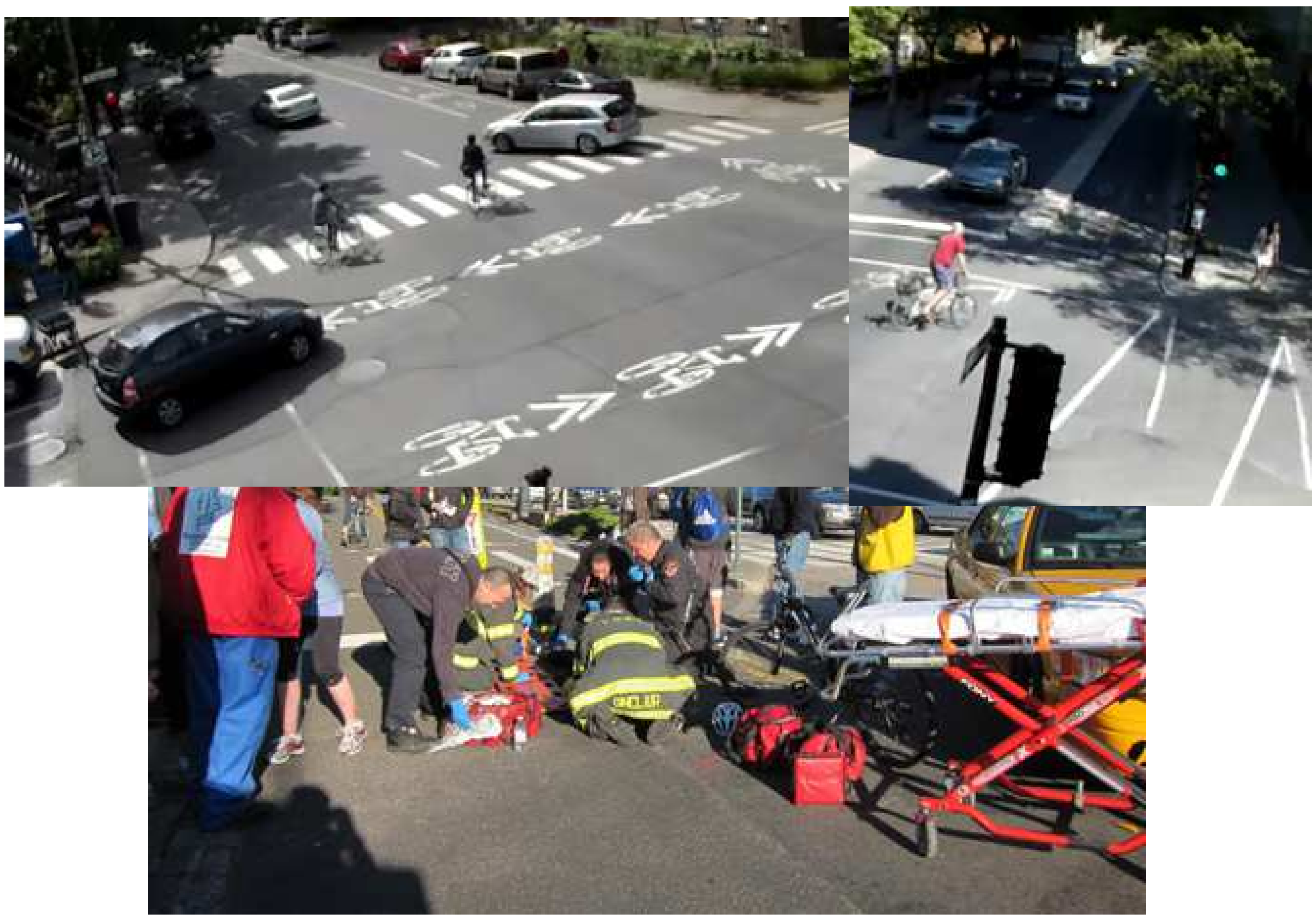


Road Safety Manual, PIARC, 2011

General definitions

- **Crashes (crashes, collisions)**: outcomes of **interactions** and sequences of actions between road users and road environment.

Interactions $\Rightarrow$ *dangerous situations* $\Rightarrow$ *crashes*



General definitions – cont.'

- **Crash occurrence (frequency):** Number of crashes occurring at a particulate site, facility or network in a period of time (this can be in one-year period or several years).

- **Crash injury severity:** Is the level of injury or property damage. Typical injury scale from police reports:

  fatal,

  major injuries,

  minor injuries,

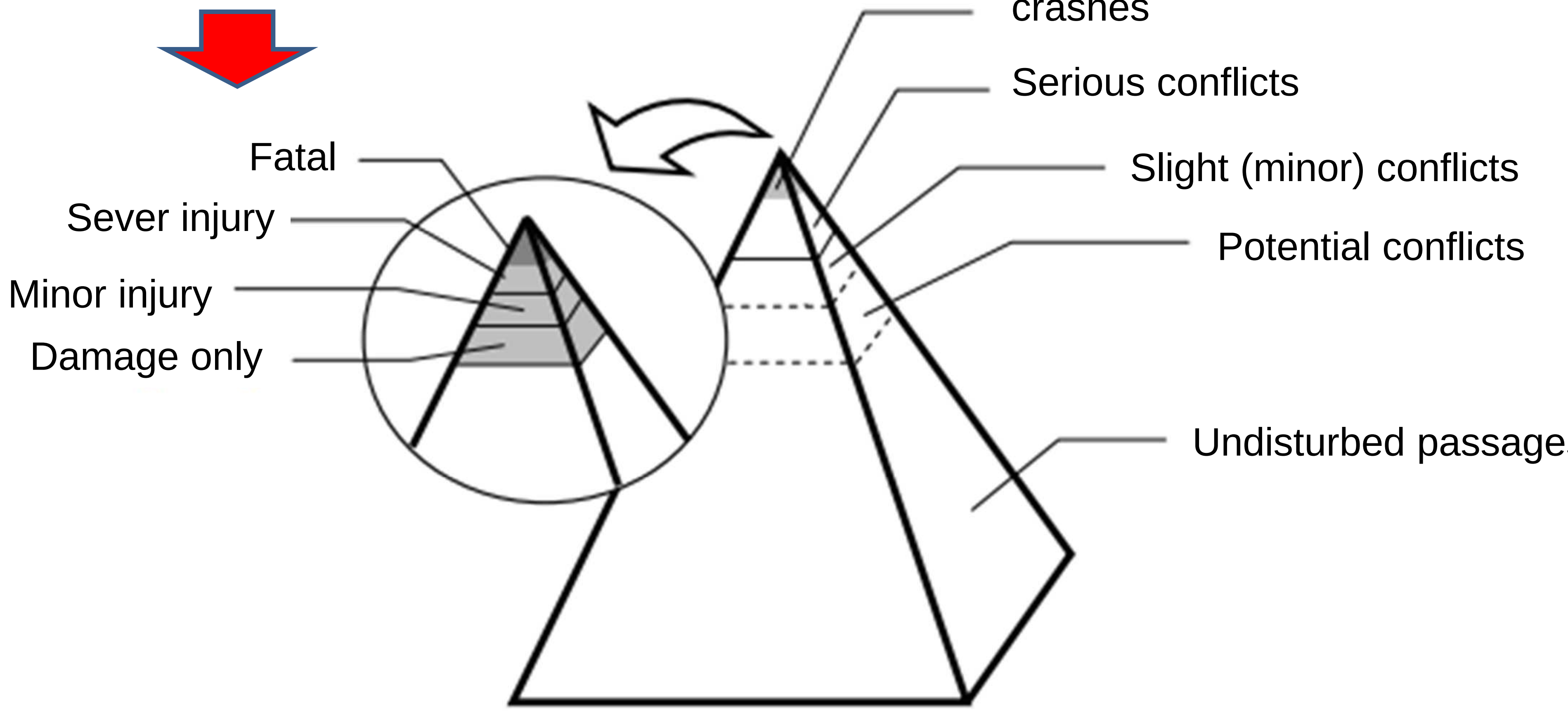
  Property damage only (PDO).

As transportation specialist, we want to target both: crash occurrence and consequences (total risk)

From undisturbed passages to crashes

Hyden's classic pyramid-model (1985)

Typical crash classification:



Alternative methods

- **Crash-based approach:**

  + *Well recognized in the literature and practice*

  - *"Crashes need to happen to be recorded"*

- **Surrogate approach (speed, conflicts):**

  This can be seen as a complementary approach. It assumes existence of a casual link to expected crash frequency

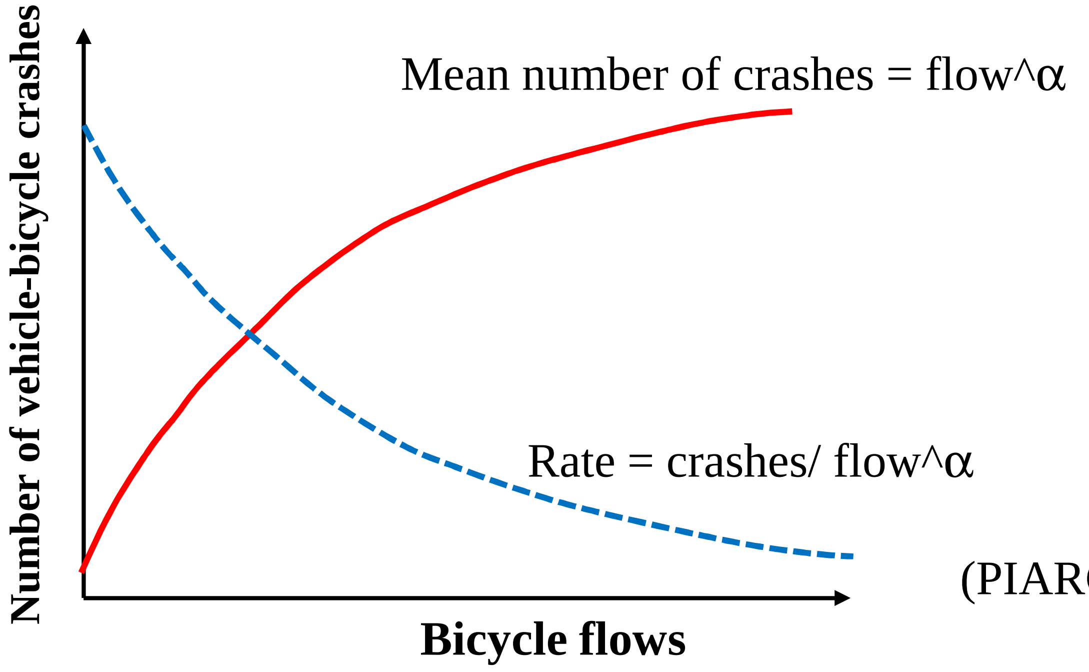
  - *Needs more validation*

  +*"we do not need to wait for crashes to happen"*

Traffic exposure and crash occurrence

Aggregate indicators:
Number of passages (traffic volume) in a given facility (e.g., intersection, midblock location) during a given period of time.

Exposure = f(number of vehicles, pedestrians and bikes)
e.g., Typical exposure indicator: Average Annual Daily Traffic AADT)



(PIARC, 2004)

Literature

❑ **Aggregate indicators of exposure:**

Previous research has focused in the relationship between vehicle-bicycle crash frequency with traffic and bike volumes *(Elvik, 2009, Strauss & Miranda-Moreno, 2012).*

$$\lambda_i = \alpha_0 Z_i^{\alpha_1} F_i^{\alpha_2}$$

λi = mean number of vehicle-cyclist collisions at a given site *i*

Zi = Bike flow at *i*

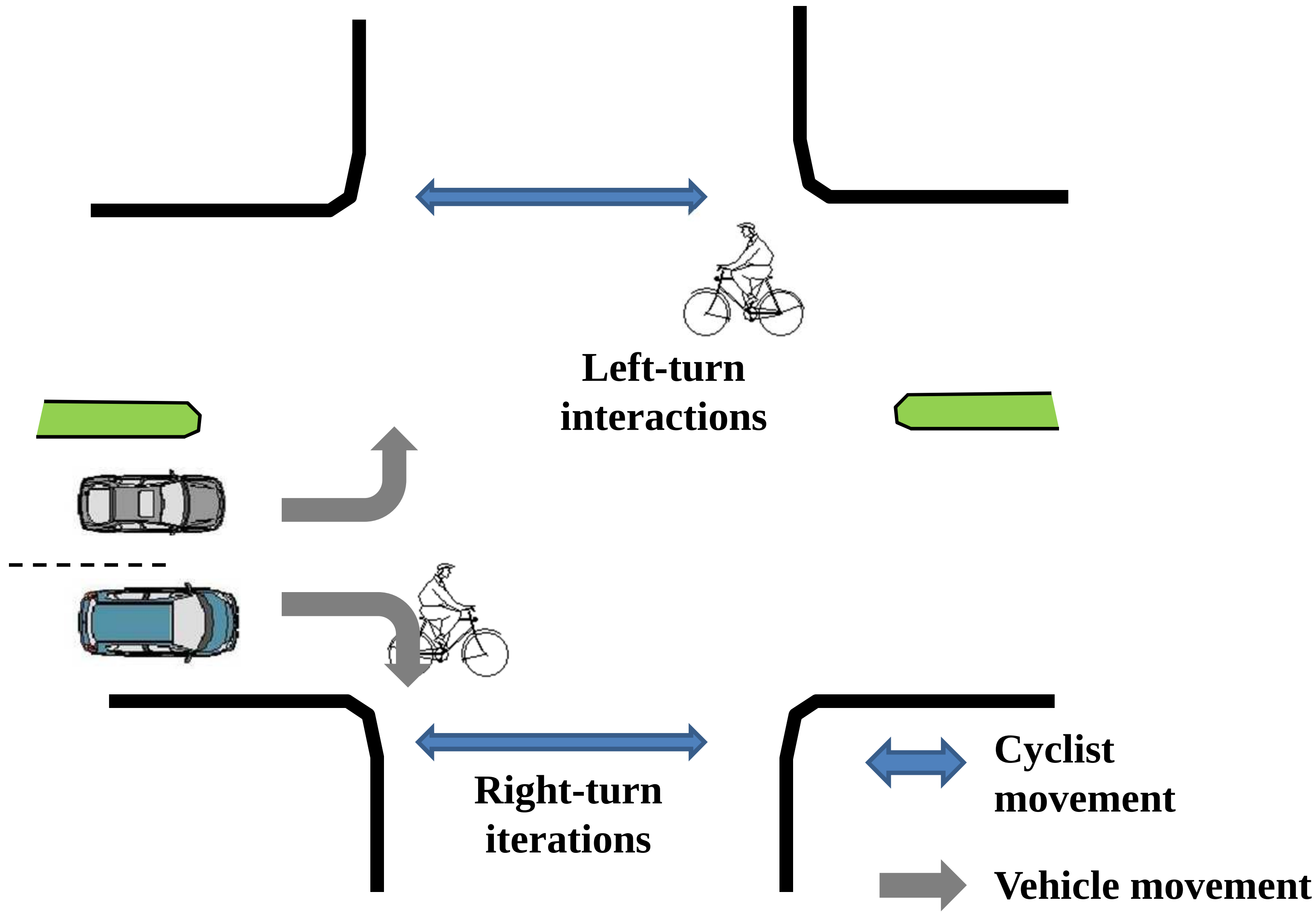
Fi = Traffic volume at *i*

Non-linear association (Elvik, 2009)

Authors	Year	Country	Study units	Sample size	Number of accidents	Type of accident	Measure of exposure	Exponent		
								Motor vehicles	Pedestrians	Cyclists
Brüde, Larsson	1993	Sweden	Junctions	285	165	Pedestrian	Entering motor vehicles, crossing pedestrians	0.50	0.72	
Brüde, Larsson	1993	Sweden	Junctions	377	432	Cyclist	Entering motor vehicles, entering cyclists	0.52		0.65
Leden, Gårder, Pulkkinen	1998	Sweden	Junctions		276	Cyclist	Entering cyclists			0.47
Leden	2002	Canada	Junctions	749	39	Pedestrian	Right turning motor vehicles, crossing pedestrians	0.86	0.48	
Leden	2002	Canada	Junctions	126	27	Pedestrian	Left turning motor vehicles, crossing pedestrians	1.19	0.33	
Lyon, Persaud	2002	Canada	Junctions	684	5280	Pedestrian	Entering motor vehicles, entering pedestrians	0.57	0.74	
Lyon, Persaud	2002	Canada	Junctions	263	1065	Pedestrian	Entering motor vehicles, entering pedestrians	0.40	0.41	
Lyon, Persaud	2002	Canada	Junctions	122	159	Pedestrian	Entering motor vehicles, entering pedestrians	0.53	0.66	
Lyon, Persaud	2002	Canada	Junctions	123	319	Pedestrian	Entering motor vehicles, entering pedestrians	0.58	0.71	
Jacobsen	2003	United States	Cities	68		Pedestrian	Share of working trips on foot		0.41	
Jacobsen	2003	United States	Cities	68		Cyclist	Share of working trips on bicycle			0.31
Jacobsen	2003	Denmark	Towns	47		Pedestrian	Kilometres walked per inhabitant per day		0.36	
Jacobsen	2003	Denmark	Towns	47		Cyclist	Kilometres cycled per inhabitant per day			0.44
Jacobsen	2003	14 European	Country	14		Cyclist	Kilometres cycled per inhabitant per day			0.58
Jacobsen	2003	8 European	Country	8		Pedestrian	Trips on foot per inhabitant per day		0.13	
Jacobsen	2003	8 European	Country	8		Cyclist	Trips on bicycle per inhabitant per day			0.48
Robinson[a]	2005	Australia	States	7		Cyclist	Kilometres cycled per inhabitant per day			0.52
Jonsson	2005	Sweden	Road sections	393	130	Pedestrian	Motor vehicle kilometres, pedestrians crossing and walking along road	0.83	0.38	
Jonsson	2005	Sweden	Road sections	393	343	Cyclist	Motor vehicle kilometres, cyclists crossing and riding along road	0.76		0.35
Geyer et al.	2006	Oakland	Junctions	247	185	Pedestrian	Entering motor vehicles, crossing pedestrians	0.16	0.61	
Harwood et al.	2008	United States	Junctions	450	728	Pedestrian	Entering motor vehicles, entering pedestrians	0.05	0.41	
Harwood et al.	2008	United States	Junctions	1433	4824	Pedestrian	Entering motor vehicles, entering pedestrians	0.40	0.45	
							Mean (simple)	0.57	0.50	0.48

Disaggregate exposure indicators

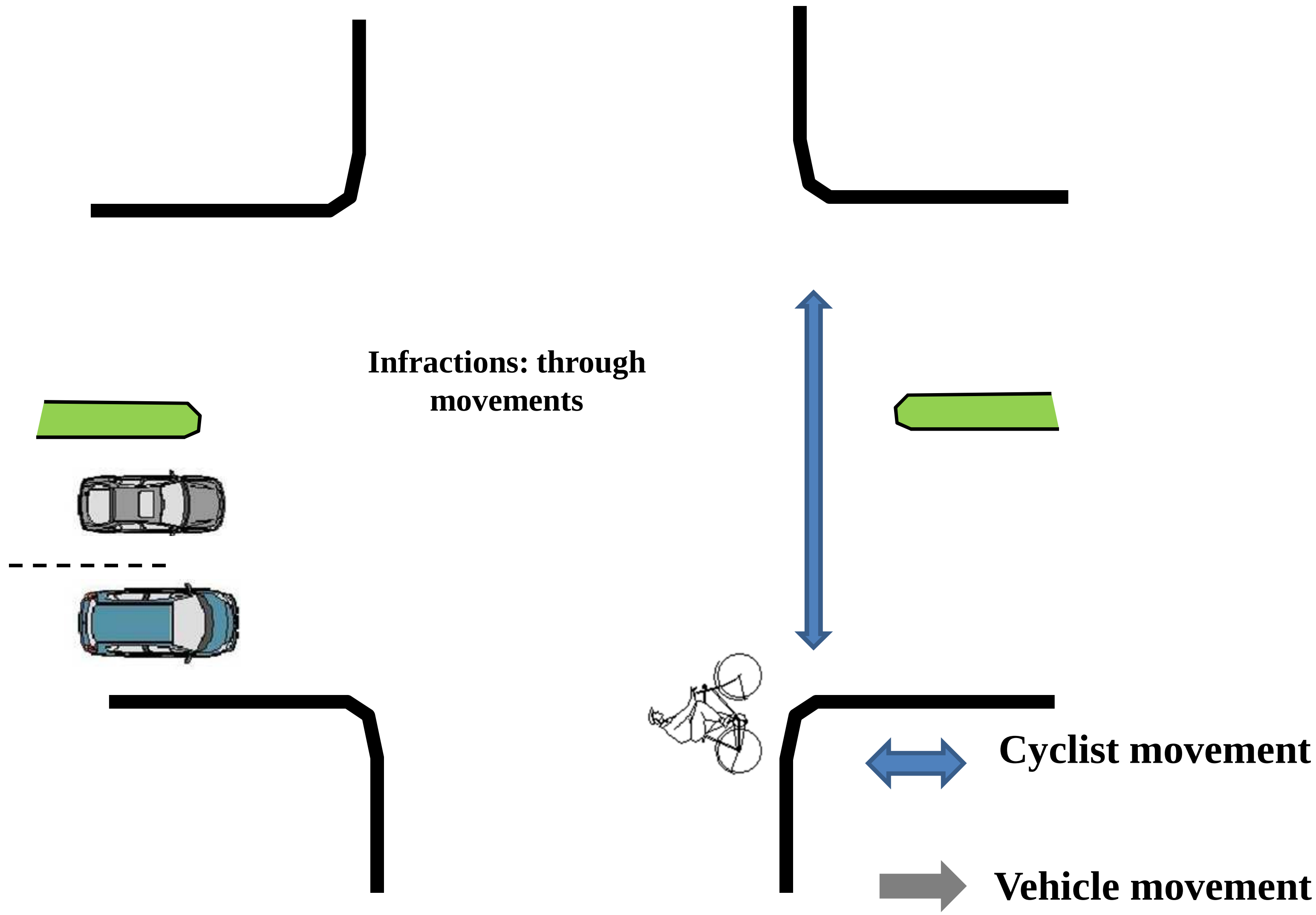
Number of interactions per movements & phase



Miranda-Moreno, Strauss (2012)

Disaggregate exposure indicator

Number of red-light crossings (violations)



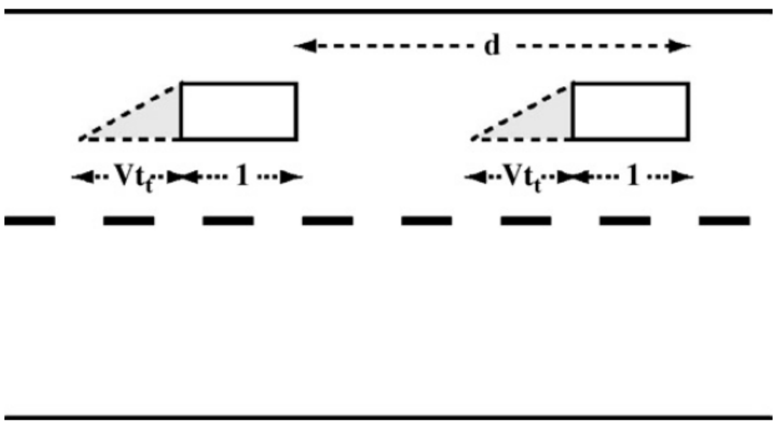
Miranda-Moreno, Strauss (2012)

Other exposure indicators

Based on the crash risk of crossing *(Routledge et al. 1976; Lassarre et al. 2007):*

- Average speed of flow
- Average traffic gap
- Average crossing time (length of crossing)
- Vehicle length

"Proportion of space that is NOT available to the pedestrian for crossing the road freely and safely".



Literature

Factors positively or negatively associated to the crash occurrence in non-motorized traffic at intersections:

- Exposure: Volume intensity, turning movements (in particular right turns)
- Road width / number of lanes / parking
- Bus stop presence / parking entrances, etc.
- Pedestrian signals
- Etc.

Miranda-Moreno, et al. 2011, Strauss et al. 2013, Morency, et al. 2012

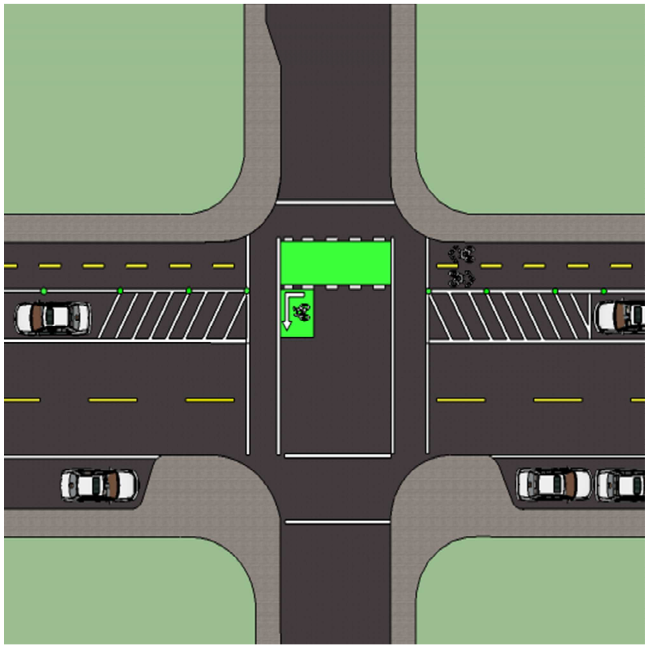
Engineering treatments

Exposure:

- ❑Separating / eliminating conflicting movements:
  - Exclusive or half phase for pedestrians/bikes
  - Restricting vehicle right-turns
- ❑Integrating ped/bike delays into traffic control design
- ❑Change intersection setting (e.g., number of approaches, roundabouts)

Engineering treatments – cont'

Traversed distance/ visibility / turning speed



- ✓ Add curb extensions (speed reduction, visibility)
- ✓ Eliminate parking in the proximity
- ✓ Add two-stage turn queue boxes
- ✓ Bike/pedestrian green light, etc.

See: NACTO Urban Bikeway Design Guide, Vélo Québec guidelines , Dutch design manual

Engineering treatments – cont’

Bike boxes

Corner Island

Cycle Track Bend Away from Street

Pedestrian platform with signal

http://wiki.coe.neu.edu/groups/nl2011transpo/wiki/bd54f/Bicycling_Facilities_in_Holland.html

The traffic "challenge"

Vehicular Capacity (c): Traffic volume (veh/h) that can pass through an intersection from a lane o group of lanes:

$$c = s \times g/C$$

Where:

c = max. hourly volume that can pass through an intersection from a lane or group of lanes, in veh/h

s = saturation flow rate in veh/h (s = 3600/sat headway)

g/C = ratio of effective green time to cycle length.

Reducing vehicle space and/or green time will affect "s" and "g"

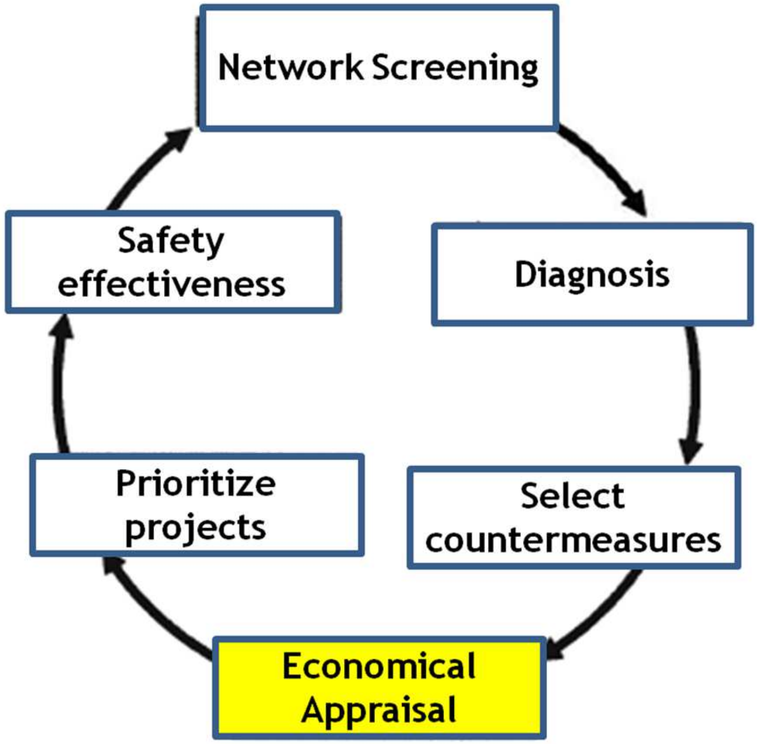
→

This can increase delays and deteriorate level of service for drivers

Fred Mannering, et al. 2010

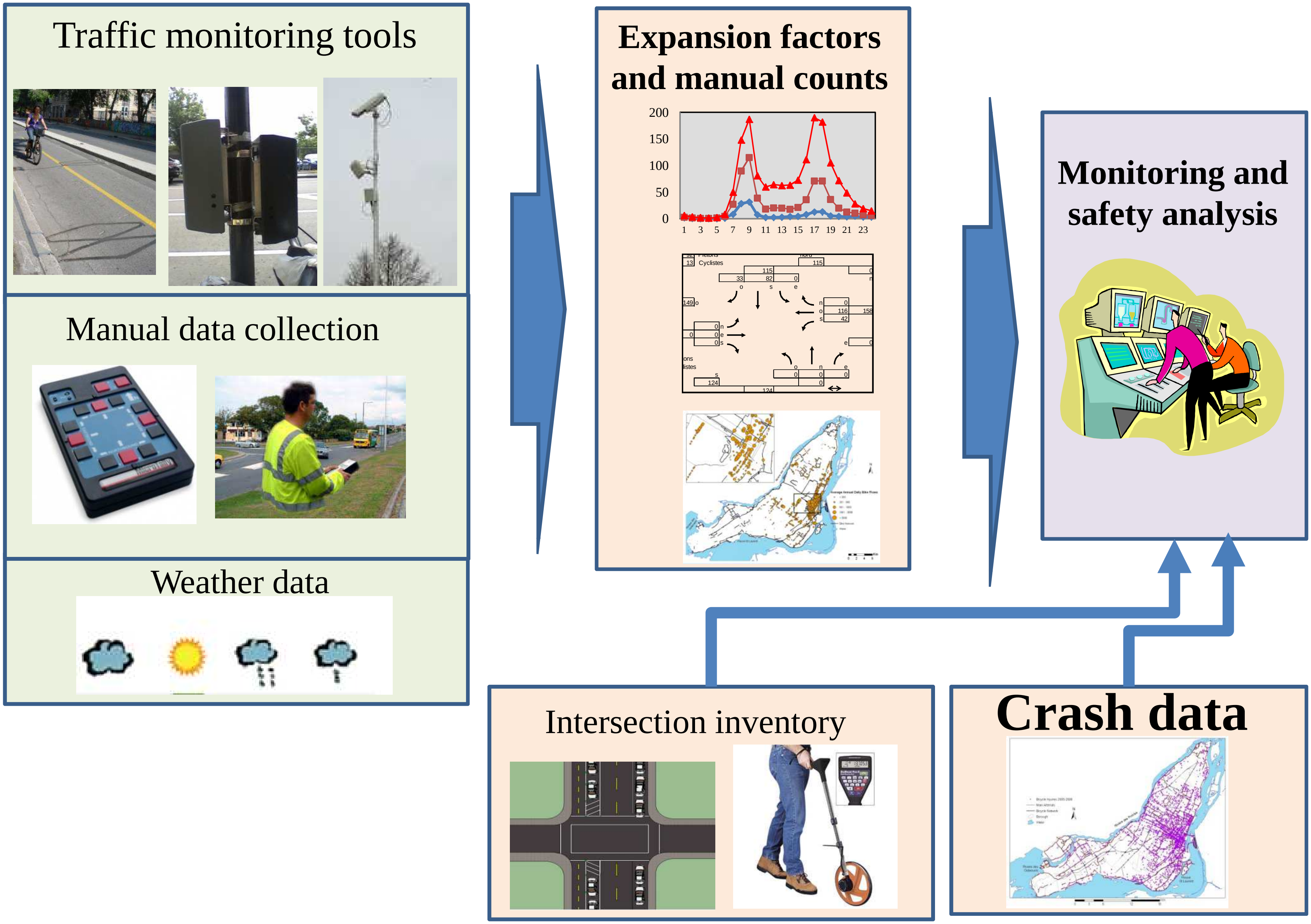
2. Data preparation and statistical analysis

Transportation management process



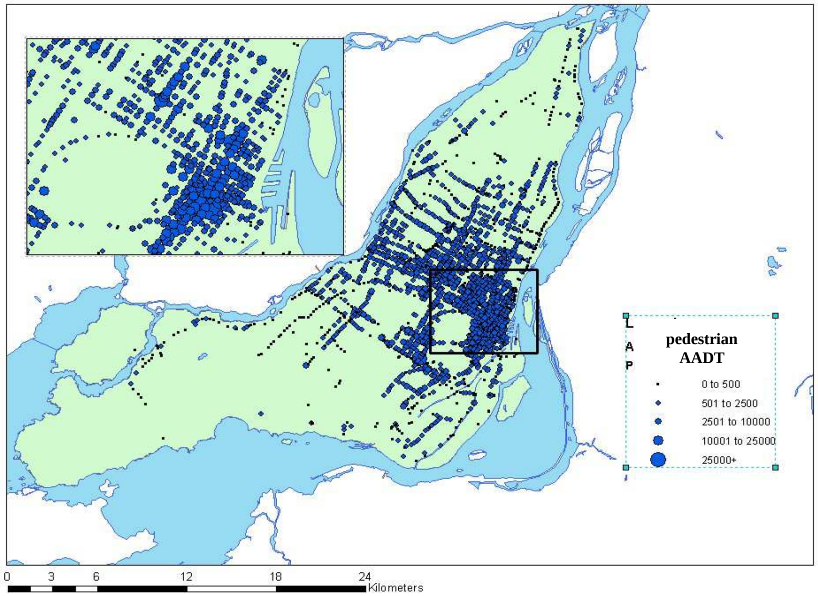
Conceptual Approaches to Road Safety
(Highway Safety Manual, AASHTO, 2010).

Data for safety analysis



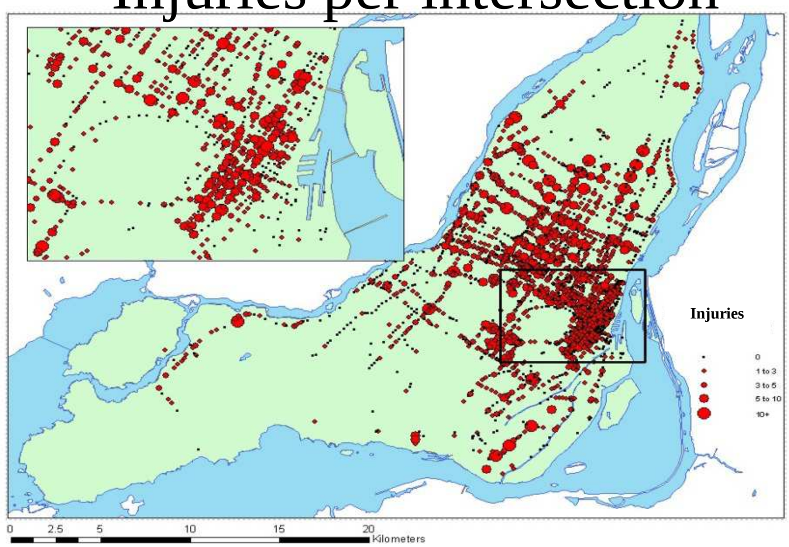
Spatial traffic and crash data

Pedestrian ADDT



Vehicular ADDT

Injuries per intersection



Traffic safety studies

Studies can be divided in two:

- Cross-sectional studies / control-case study
- Before-after observational studies

Applications:

- Safety effectiveness: before-after studies
- Network screening: mapping crash risk

Methods for crash risk estimation:

- Raw risk approach
- Model-based approach

“Raw” crash-based method

$$R_i = \frac{y_i \times 10^6}{365(T_i \times AADT_i)}$$

R_i = injury rate of intersection i (injuries per million of cyclists per unit of time)

y_i = bicycle injury frequency at intersection i during T_i

T_i = period of analysis (years)

$AADT_i$ = Average annual daily bicycle volume of intersection *i*

Road Safety Manual, PIARC, 2011

Example 1: Control-case study on bike lanes

Evaluate the cyclist risk (injury rates) in the following 3 street sections with bicycle lanes. Compare the risk with the similar 3 street sections without bicycle lanes. Determine the safety effectiveness of bike lanes.

	ID	**Section**	**Years**	**L. (km)**	**AADT**	**Number of injuries**
Facilities	1	Milton	1.5	0.52	1,797	2.0
	2	Hutchison (south)	2.5	0.78	2,343	2.0
	3	Prince Arthur	3.5	0.96	805	3.0
Controls	1	Control 1(Hutchison north)	2.5	0.75	1,958	3.0
	2	Control 2 (Brebeuf)	3.5	1.1	1801	7.0
	3	Control 3 (Boyer)	3.5	1.5	803	7.0

Solution

$$R_i = \frac{y_i \times 10^6}{365(T_i \times L_i \times AADT_i)}$$

Injuries per million of bike-km

	ID	Section	Years	L. (km)	AADT	Number of injuries	Injury Rates	Average values
Facilities	1	Milton	1.5	0.52	1797	2.0	3.91	
	2	Hutchison (south)	2.5	0.78	2343	2.0	1.20	
	3	Prince Arthur	3.5	0.96	805	3.0	3.04	2.72
Controls	1	Control 1(Hutchison north)	2.5	0.75	1958	3.0	2.24	
	2	Control 2 (Brebeuf)	3.5	1.1	1801	7.0	2.77	
	3	Control 3 (Boyer)	3.5	1.5	803	7.0	4.55	3.18

Average reduction = [1 – (3.18/2.72)]*100 = - 17%

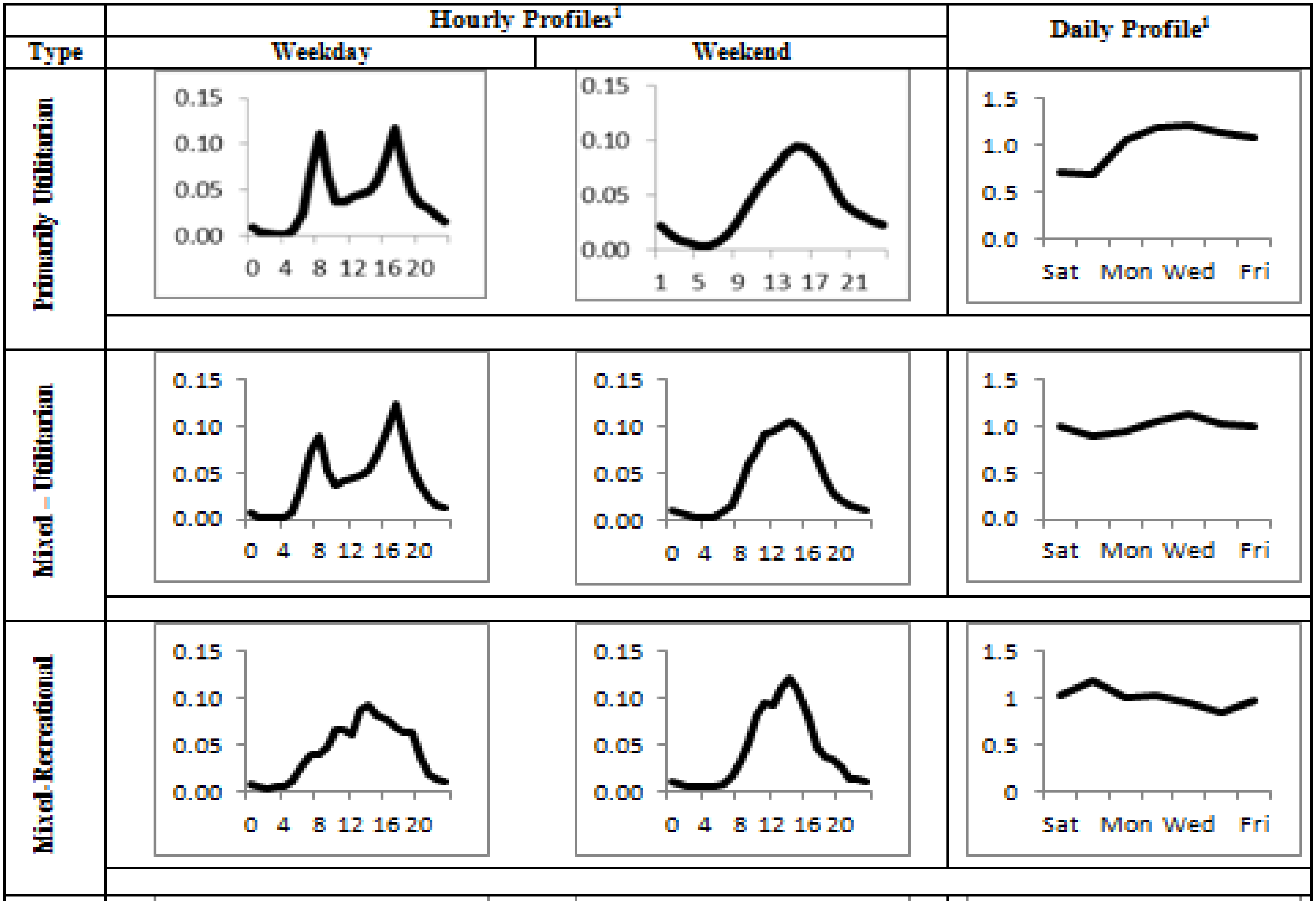
Elements to take into account:

- **Average Annual Daily Traffic (ADDT):** How to determine this using short-term counts considering temporal/weather effect?

- **Crash (injuries) data:** How many years of data? Low-mean problem?

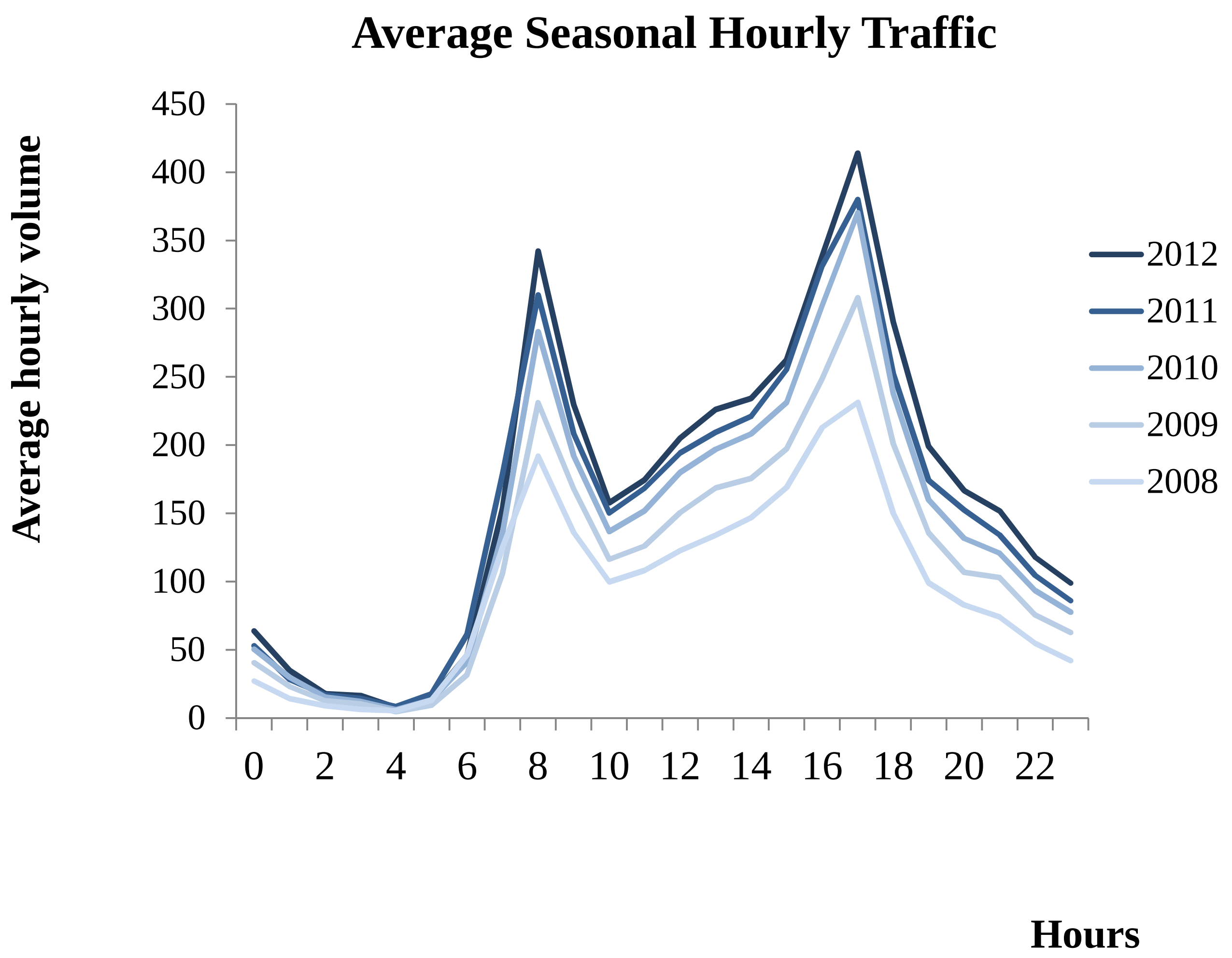
- **Temporal changes on AADT and y_i:** E.g., before and after the installation of a bike infrastructure.

Adjustment factors for different traffic patters



Miranda-Moreno, Nosal, 2012

Volume trends in bicycle facilities in Montreal

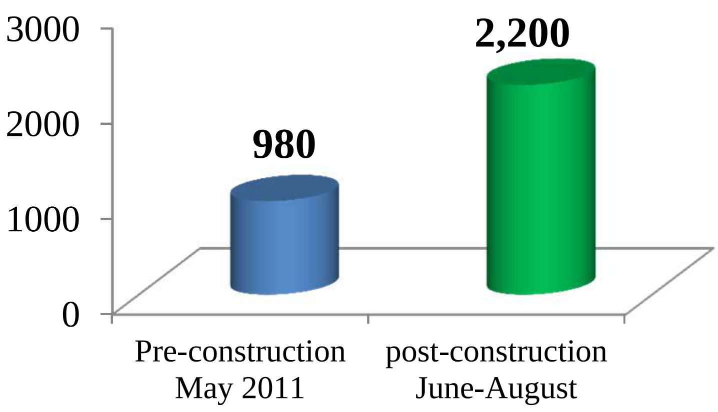


Before-after traffic volumes

1) Bicycle lanes, Ave. Laurier, Montreal

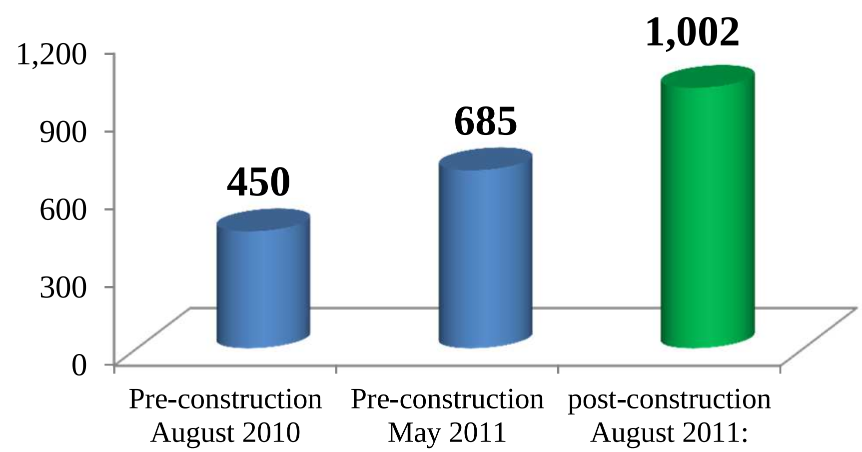
2) Cycle track, Ave. Laurier, Ottawa

Standardized daily flows



Increase: > 100% with respect to May 2011

Standardized daily flows



Increase: ~ 50% with respect to May 2011

Example 2: Bicycle traffic

To improve the traffic safety of cyclists, the city of Ottawa recently install a bicycle facility on Ave Laurier – a bidirectional cycle track. The following data was collected before and after the installation in 2011:

- April 29 (Fri): 150 bikes from 8:00 to 10:00am
- May 11 (Wed): 95 bikes from 9:00 to 10:00am
- May 16 (Mon): 211 bikes from 8:00 to 10:00am
- June 15 (Wed): 284 bikes from 8:00 to 10:00am
- June 30 (Fri): 125 from 9:00 to 10:00am
- July 25 (Mon): 299 from 8:00 to 10:00

Weather conditions: no rain, temp between 15°C and 25 °C

Question: The facility started operating on June 06, 2011. Determine the increase or decrease in terms of AADT according to the manual counts (adjusting only for temporal trends).

Use the expansion factors obtained from a counting stations located in a bike facility with similar traffic patterns in the same city.

Expansion factors from automatic counts

Hourly traffic variations (in percentage out of 100)

Hour	Weekday	Saturday	Sunday
0	1.03	3.08	3.26
1	0.48	1.84	2.07
2	0.24	1.10	1.26
3	0.20	0.93	1.12
4	0.18	0.43	0.48
5	0.58	0.40	0.39
6	1.91	0.68	0.64
7	5.72	1.26	1.14
8	11.24	2.46	2.10
9	7.07	3.66	3.31
10	4.18	4.64	4.55
11	4.22	5.80	6.03
12	4.77	6.75	7.27
13	5.11	7.71	8.74
14	5.11	8.49	9.56
15	5.80	8.49	9.41
16	8.01	8.15	8.58
17	10.71	7.40	7.56
18	7.23	5.89	5.84
19	4.78	4.96	4.42
20	3.75	4.08	3.81
21	3.35	4.13	3.30
22	2.44	3.88	2.79
23	1.89	3.79	2.39

Day	Daily expansion factors
Sunday	0.68
Monday	1.03
Tuesday	1.16
Wednesday	1.19
Thursday	1.12
Friday	1.08
Saturday	0.74

Month	Monthly expansion factors
April	0.80
May	1.09
June	1.17
July	1.27
August	1.20
September	1.13
October	0.67
November	0.55

Winter months factors are 0

Solution

$$ADDT = V \times [(1/Fh) \times (1/F_D) \times (1/F_M)]$$

Fh, F_D, F_M - hourly, daily and monthly expansion factors

Hour	Day / month	Volume	Hourly factors	Daily factors	Monthly factors	Daily average (corrected)
8 -10	Fri / April	150	0.1831	1.08	0.8	948
9 -10	Wed / May	95	0.0707	1.19	1.09	1036
8 - 10	Mon / May	211	0.1831	1.03	1.09	1026
					Average daily (before):	1004
After						
8-10	Wed / June	284	0.1831	1.19	1.17	1114
9 - 10	Fri / June	125	0.0707	1.08	1.17	1399
8 -10	Mon / July	299	0.1831	1.03	1.27	1248
					Average daily (after):	1254

Estimated % increase: 0.25

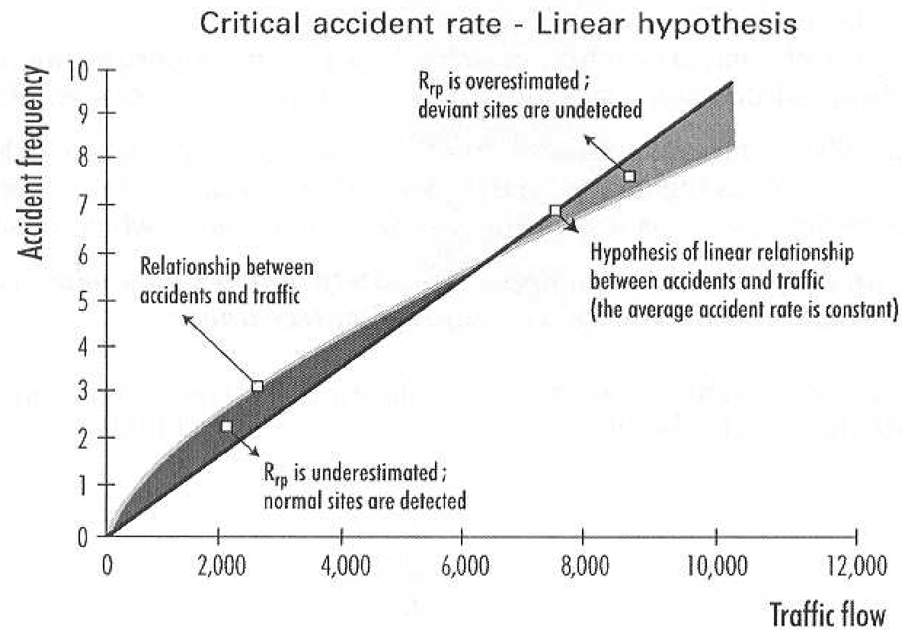
Shortcoming of "raw crash" risk indicator

a) Uncertainty is not taken into account

b) They do not consider the possibility of a **non-linear relationship** between crashes and traffic exposure

c) Variations in **roadway characteristics:** raw estimates ignore the effect of site-specific attributes (e.g., geometry and signalization factors)

So… it can lead to wrong estimations



Road Safety Manual, PIARC, 2011

2.1 Statistical models for crash frequency analysis

Poisson Approximation

Since a (crash) event has a very low probability of occurrence and a large number of trials exist (e.g. million entering vehicles, vehicle-miles-traveled, etc.), the binomial distribution is approximated by a Poisson distribution.

$$P(Y=n)=\binom{N}{n}\left(\frac{\mu}{N}\right)^{n}\left(1-\frac{\mu}{N}\right)^{N-n}\cong\frac{\mu^{n}}{n!}e^{-\mu}$$

Where,

n = 0, 1, 2, … , N (the number of successes or crashes)

μ = the mean of a Poisson distribution

Negative Binomial (Poisson-Gamma) model

$$Y_i \mid \theta_i \sim Poisson(T_i \theta_i)$$

$$\sim Poisson(T_i \mu_i e^{\varepsilon_i})$$

Random effect

$e^{\varepsilon_i} \sim Gamma(\phi, \phi)$

where: Y_i = number of crashes at site i

Ti = observation time at i

θ_i = mean number of crashes at i

$$\mu_i = \beta_0 F_{1i}^{\beta_1} F_{2i}^{\beta_2} \exp(\beta_3 x_3 + \ldots + \beta_k x_k)$$

F_{1i}, F_{2i} = Traffic flows,

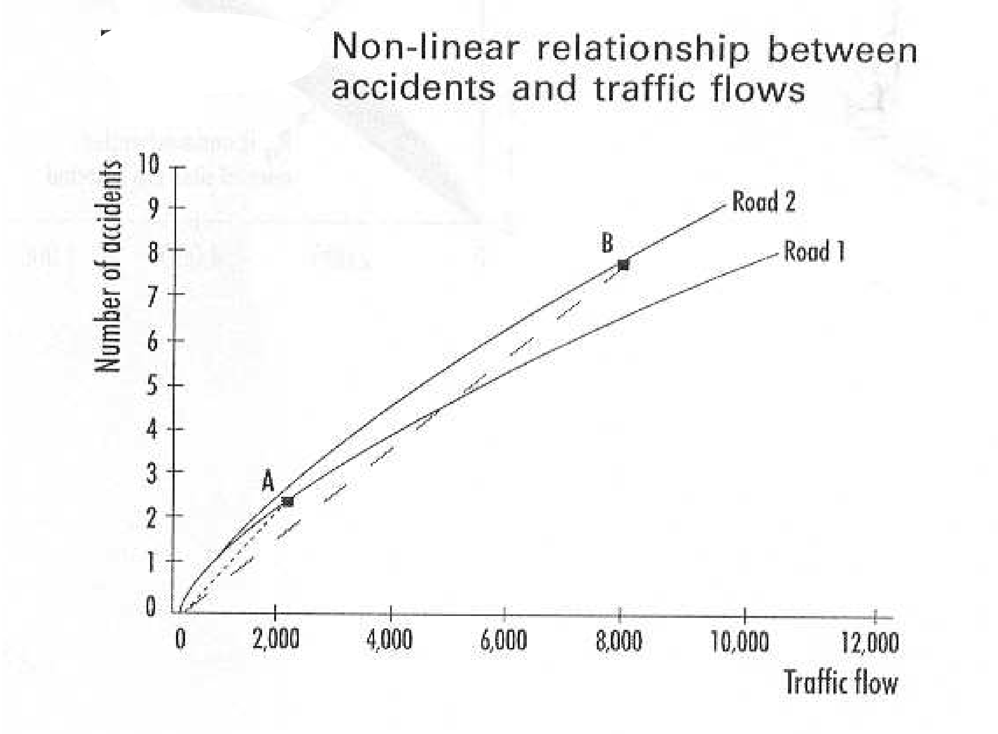
β_0, β_1, …, β_k = regression parameters

xk = geometry/traffic control attributes

Typical safety performance functions (SPF)

$$\mu_i = F_{1i}^{\beta_1} F_{2i}^{\beta_2} \exp(\beta_0 + \beta_3 X_{3i} + \ldots + \beta_k X_{ki})$$ (for intersections)

$$\mu_i = L_i (F_{1i} + F_{2i})^{\beta_1} \exp(\beta_0 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})$$ (for sections)



(PIARC, 2004)

Posterior analysis using NB model

The Poisson/Gamma or Negative Binomial model can be written as (considering T constant):

$$\mathrm{Y_i} \mid \theta_\mathrm{i} \sim \mathrm{Poisson}(\theta_\mathrm{i})$$

$$\theta_\mathrm{i} \sim \mathrm{Gamma}(\phi, \phi / \mu_\mathrm{i})$$

Applying Bayes' theorem:

$$p(\theta_i \mid y_i) \propto f(y_i \mid \theta_i)\pi(\theta_i)$$

$$\propto \frac{e^{-\theta_\mathrm{i}}(\theta_i)^{y_\mathrm{i}}}{y!} \cdot \frac{(\phi / \mu_\mathrm{i})^{\phi}}{\Gamma(\phi)} \theta_i^{\phi-1} e^{-(\phi/\mu_\mathrm{i})\theta_\mathrm{i}}$$

$$or \quad \theta_i \mid y_i \sim Gamma(y_i + \phi,\ 1 + \phi / \mu_i)$$

EB estimator based on NB model

Knowing that $p(\theta_i|y_i)$ is a Gamma distribution with shape $a=(y_i+\phi)$ and scale parameter $b=(1 + \phi/\mu_i)$, the posterior mean and variance of θ_i is given by:

$$E(\theta_i \mid y_i) = \frac{y_i + \phi}{1 + \phi / \mu_i} \quad \textit{and} \quad \mathrm{Var}(\theta_i \mid y_i) = \frac{y_i + \phi}{(1 + \phi / \mu_i)^2}$$

$$EB_\theta = E(\theta_i \mid y_i) = (1 - w_i) y_i + w_i \mu_i$$

$$\text{with} \quad w_i = \phi / (\phi + \mu_i)$$

This is known in the safety literature as the EB estimator and has been widely implemented by researchers and government agencies for safety analysis.

Example 3: bicycle injury analysis at signalized intersections

Carry on the following tasks using a sample of signaled intersections.

- Define risk exposure measures according to the definitions provided before.
- Develop collision occurrence models and select the most appropriate ones (report only the best models).
- Identify the main contributing factors (traffic conditions, geometric, controls, etc.) using the parameters, confidence intervals, elasticities.

Data description :

Collision data

- Bike injuries: represents the traffic-related bicycle injuries for a six-year period and includes all vehicle-bicycle collisions within a 15-meter radius.
- AADT_bike: Average annual daily traffic for bikes
- AADT_vehicle: Average annual daily traffic for motor vehicles.
- AADTs were obtained using 8-hour manual counts. They have been expanded using appropriate factors (temporal and weather factors for bikes).
- Geometry variables

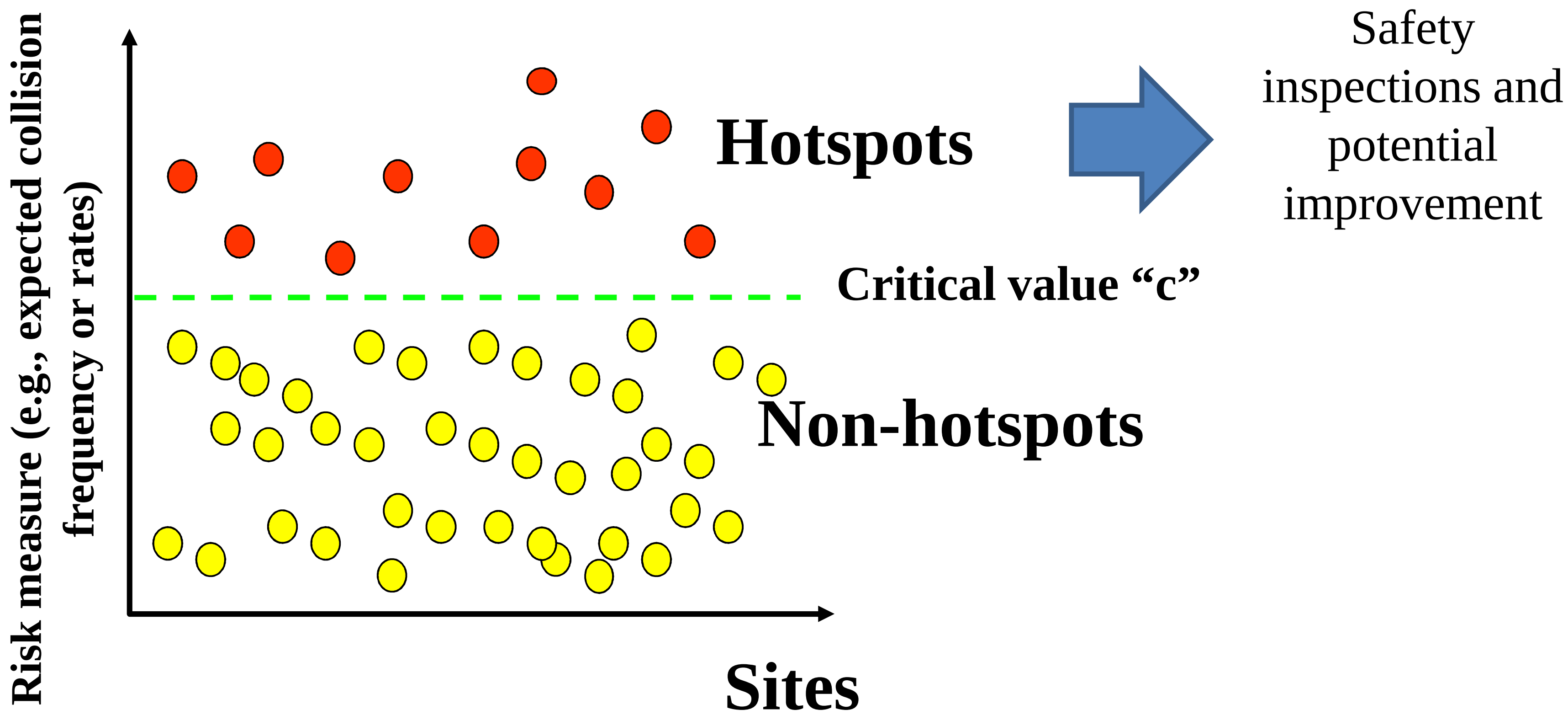
2.2 Before-After Studies

Before-after studies

- **Types of methods**
  - Naïve before-after method
  - Before-after method with control group
  - Empirical Bayes approach (with control group)
- **Important Issues**
  - Selection bias
  - Regression-to-the-mean
  - Non-linear relationship between crashes and traffic flow
  - Low mean

Site Selection Bias

Sites that are treated are not randomly selected. Therefore, treatment effectiveness cannot be estimated directly without taking this into account.



Regression-to-the-mean

- This consists of the general tendency of **extreme values** to regress to mean values.
- Then, if one treats hazardous locations with high observed or estimated crash history, the effect **RTM** should be considered.

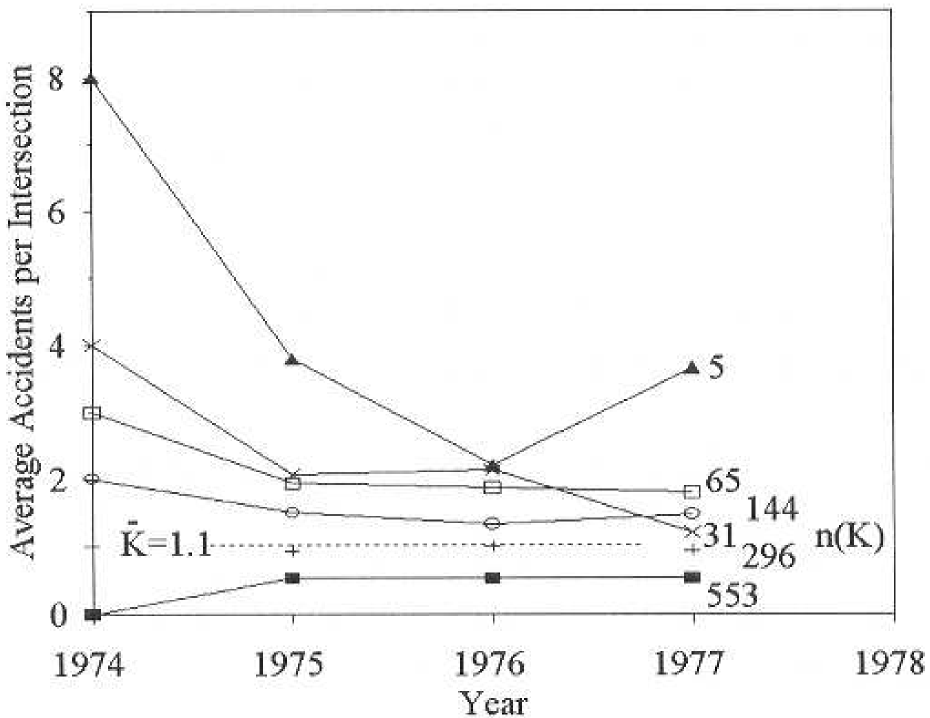


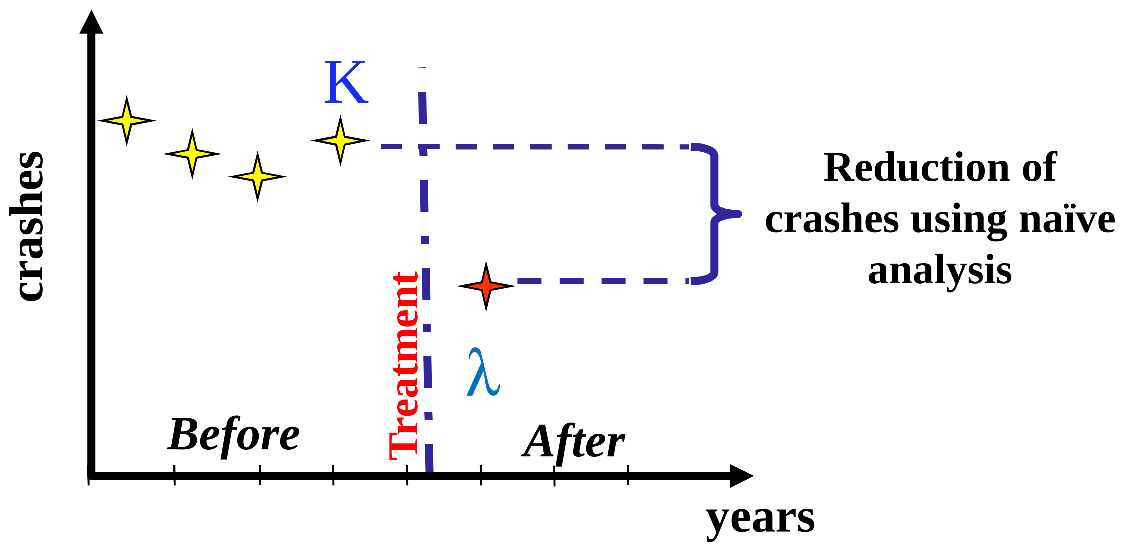
Figure 11.1. How accident counts regress to the mean.

Hauer, 1997

Before-after methods for evaluation of treatment effects

1. Naïve before-after approach
2. Before-after approach using a control group
3. Empirical Bayes approach using a control group

Naïve before-after approach



- The change in crashes for a ***treated intersection*** is given by:

$$\text{Treatment effect } (\delta) = K - \lambda$$

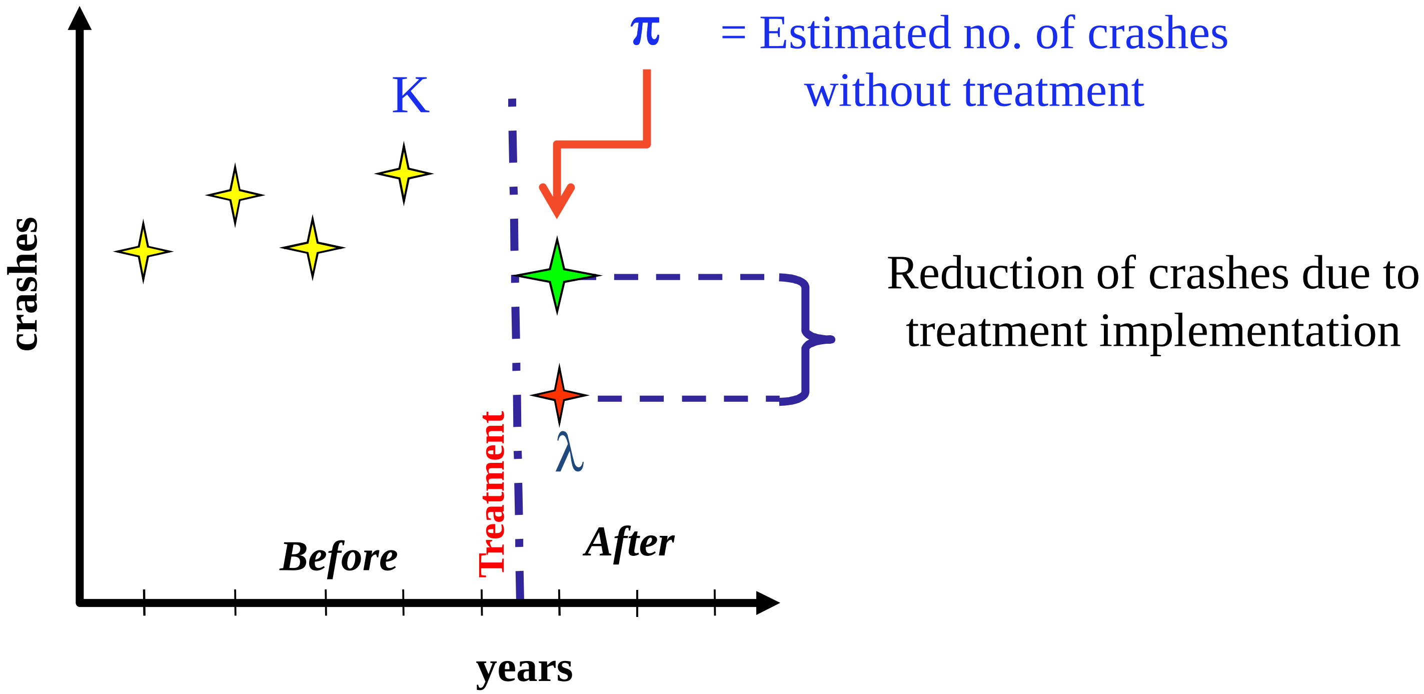
where:

K = # of crashes in the before period

λ = # of crashes in the after period

Assuming the observation period (time) before and after the treatment is the same

Before-after study using a control group



π is estimated using a **site or a group of similar sites in which non-treatment** have been applied

Treatment effect (δ) = π - λ

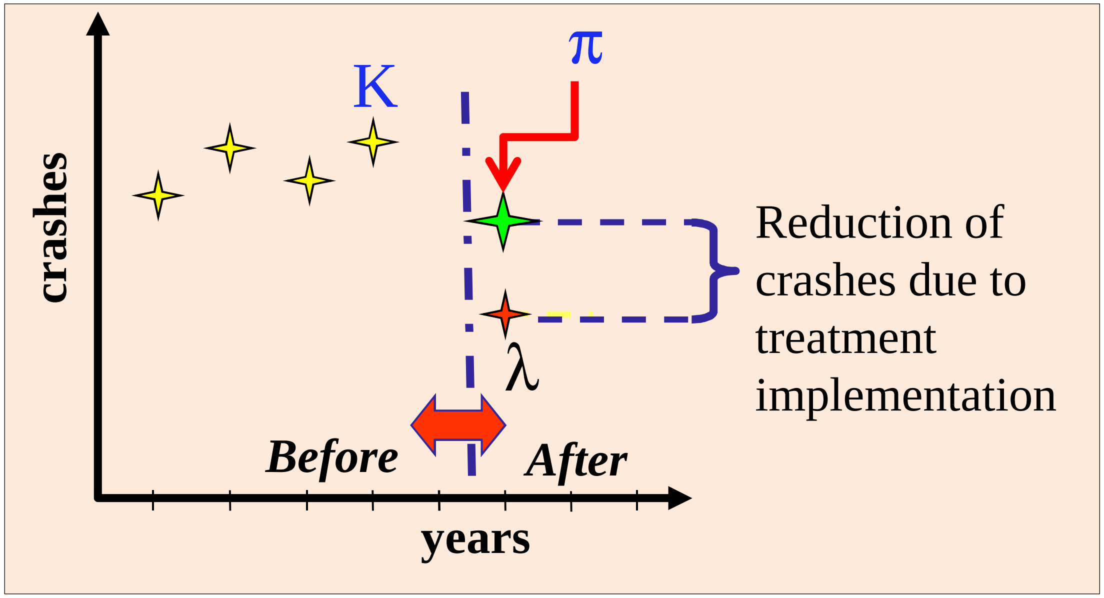
Challenge: how to select an appropriate control group

Before-after study using a control group

$$\pi = r_t \times K$$

r_t = ratio between the no. of crashes in the after / no. of crashes in the before period on the comparison site (or group)

π = Estimated no. of crashes without treatment



Assumption: factors affecting safety changed in the same way on both treatment and comparison group

Example 4: Before-After study using a control group

A “Bike box” was installed to increase visibility of cyclists and reduce conflicts between motor vehicles and cyclists, particularly in potential “right-hook” situations. Injuries in the treated and control intersections have been recorded in the 3 years before and after the installation:

C. Whitney, et al.

	Before period (2002-2004)		After period (2005-2007)	
Intersection	AADT	# of crashes	AADT	# of crashes
Treated site	925	22	1,259	17
Control site	825	17	942	16

Solution:

Calculating injury rates

	K	λ
Treated intersection	5.1	3.6
Control intersection	5.2	4.2

r_t = Injury rate in the after /injury rate in the before period (comparison site) = 4.2/5.2 = 0.8

$\pi = K * r_t = 5.1 * 0.8 = 4.1$

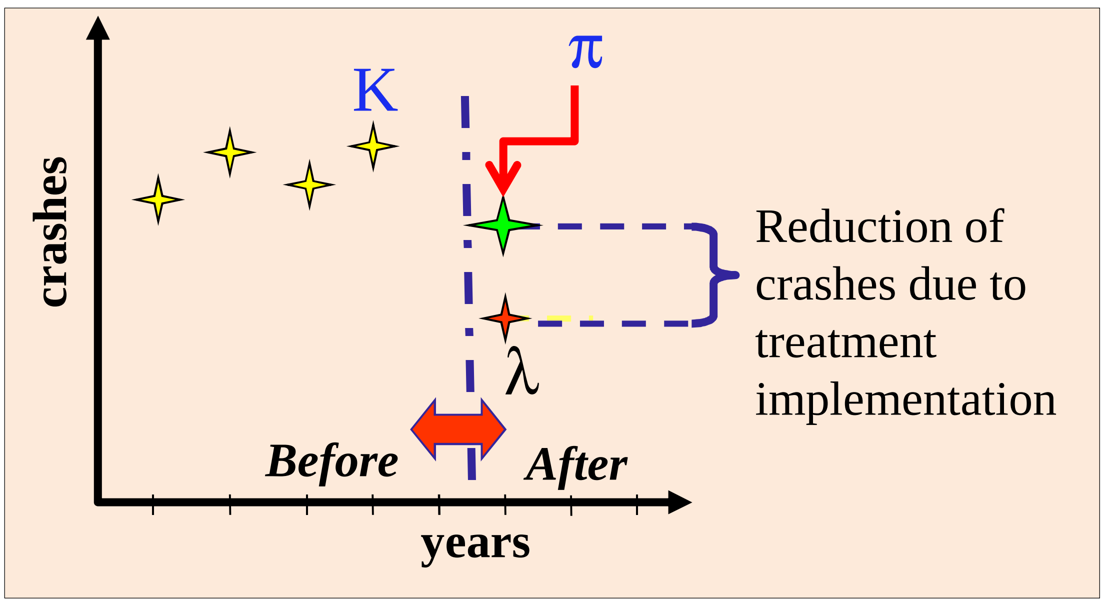
Estimated crash reduction = 4.1 – 3.6 = 0.5 injuries

Before-after study using EB approach

$$\pi = EB \times r_t$$

r_t = ratio between the no. of crashes in the after / no. of crashes in the before period on the comparison site (or group)

EB = Expected number of "before" crashes on the treated site (or group) using the posterior distribution of crashes



Final remarks

- ❑ Keep in mind the quality of data:

- Not all reportable crashes are actually reported
- Crash location can be inaccurate
- Bike/ped demand is very sensitive to land-use, weather, etc.

- ❑ In non-motorized safety: low mean can be an issue

- ❑ Surrogate analysis could be a good complement: more validation is needed.

- ❑ A lot of opportunities for research