Chapter 2: Research Methods

Previous studies of bicycle crash incidents fall into three general categories: analysis of crash rates by the number of bicycle trips taken (or number of cyclists) within a particular area or population; analysis of crash rates by distance travelled by individual cyclists (or particular groups of cyclists); and crash-type frequency analysis. These categories complement one another, and each has its strengths and limitations. Cycling safety research is hampered by a general lack of detailed empirical information on levels of cycling activity. Most traffic counting machines do not detect bicycles reliably. Large-scale manual traffic counts are more informative, but are costly, so they are usually run for only one day per station. If a count is conducted during poor weather, it is likely to underestimate the usual level of bicycling activity. In many bicycle crash studies, therefore, the subjects' degree of exposure to crash risk can only be estimated.[15]

Measuring crash rates requires both collision data and exposure estimates, which are usually gathered through surveys. A survey by Lisa Aultman-Hall[16] collected both types of information. Respondents were asked to provide information on the distances they cycle, and to report all the crashes they experienced. Studies of this kind can provide much-needed information about the relative safety of different routes and cycling facilities. For Aultman-Hall's sample of 1,359 Toronto commuter cyclists, most crashes (collisions and falls) occurred on roadways, since that is where most cycling activity took place. However, crash rates *per kilometre* were significantly higher on sidewalks and multi-use paths. The study also found that crash rates were dependent on the riding habits of the cyclists, such as the weekly cycling distance and the type of cycling facilities preferred. Those who rode more exhibited lower crash rates. Cyclists who reported that they often ride on sidewalks exhibited higher crash rates, even on roads, than those who said they never ride on the sidewalk. This kind of information can help determine the type of safety messages that should be provided to all cyclists, and can also identify opportunities for infrastructure improvements.

---

[15] "Exposure" is a function of the number of trips and the average trip length. Cyclists who make frequent and/or long trips are said to have a greater exposure to traffic hazards than those who cycle less.

[16] Doherty *et al*, 2000

Distance-based crash rates (or injury and/or fatality rates) for cyclists are sometimes compared with rates for other modes, to give an idea of their relative safety. However, since bicycle trips are usually shorter than auto trips, on average, it is perhaps more appropriate to compare them on a per-trip basis. These rates can also provide an indication of the relative safety of cycling in different places. For instance, the number of cyclist fatalities per 100 million bicycle trips in different countries ranges from lows of 1.6 in the Netherlands and 2.4 in Germany, up to 26 in the US.[17] Based on traffic fatality records, and travel estimates from the City's most recent large-scale cycling survey,[18] the cyclist fatality rate for the City of Toronto works out to between 6 and 10 fatalities per 100 million bicycle trips. The accuracy of such calculations is limited by a reliance on self-reported information — the personal accounts of cycling activity collected by surveys.

One of the benefits of this kind of survey-based crash research is that it captures information on all kinds of cycling accidents, including those that may not show up in police and hospital records. This provides a way of estimating the proportion of crashes that are not reported to authorities, so that the full spectrum of cycling accidents can be better understood. It also provides a sense of the relative significance of different kinds of crashes. Most surveys find that collisions typically result in more serious injuries, but that falls are more common.

While such studies can capture a representative sample of the kinds of *incidents* affecting the cyclists surveyed, it is difficult to obtain a sample that is truly representative of the local cyclist *population*. Other studies have focused on specific locations, rather than on particular groups of cyclists. For instance, Wachtel and Lewiston counted the number of cyclists using the road versus the sidewalk, on three arterial roads in Palo Alto, California.[19] By comparing these numbers with police collision reports, they arrived at a measure of the relative safety of these different facilities. Strictly speaking, the findings apply only to the specific locations studied. More general results would require large amounts of data, which would take a long time to collect, given the relatively low levels of bicycle use in most North American cities. Studying conflicts ("close-calls") rather than actual crashes can yield similar information more quickly.

---

[17] Pucher and Dijkstra. 2000

[18] Decima, 2000

[19] Wachtel. and Lewiston, 1994

The majority of crash research done in the past has been of the third type, crash-type frequency analysis, of which this study is an example. One reason it is the most common is that it uses pre-existing databases (police accident reports and hospital records) which are usually quite detailed and extensive. However, these sources of data capture information on only a fraction of all crash incidents. Falls and minor collisions are rarely reported to police, and only those resulting in injuries appear in hospital records. Also, while collision frequency analysis can highlight sites where crashes are unusually frequent, it cannot tell us whether these are hazardous locations or simply heavily-used bicycle routes. Despite these limitations, such studies do capture a representative sample of the cyclists involved in collisions. That is, unlike surveys, there is no mechanism that might selectively discriminate against certain portions of local cycling population appearing in collision reports or hospital records.[20] Thus, the findings of collision frequency analysis provide reliable, detailed information that can be generalised with confidence. Furthermore, since collisions generally result in more serious injuries than do falls, and because the fear of collisions acts as a significant deterrent to cycling, they are of great interest to organisations that implement public safety initiatives, and promote cycling.

In a typical collision frequency analysis, a large number of collision reports are studied and classified according to a “crash typology,” a system of categories, each of which defines a specific sequence of events that best describes a particular type of collision. The process generates summary statistics for the different collision types, and the cyclists and drivers involved. It allows investigators to uncover some of the factors that may have contributed to the occurrence of collisions, factors that may only stand out when the data is aggregated. The findings are useful in designing public awareness campaigns, driver- and cyclist-education materials, and traffic regulation enforcement programs. With enough data, collision frequency analysis can also identify “hot spots” where collisions occur unusually often, suggesting the need for operational or infrastructure modifications. It can also reveal areas where the reporting process and police officers’ awareness of bicycle traffic safety issues might be improved. Addressing these concerns would help to improve the collection of basic data on bicycle collisions, a need that the Toronto Coroner identified after completing his study. It could also help improve relations between police and cyclists.

[20] Stutts and Hunter, 1998

2.1 Data Collection and Organisation

The Toronto Bicycle/Motor-Vehicle Collision Study consisted of three main stages: collection and organisation of the information, analysis of the data, and presentation of the analysis results in graphical format. The methods are based on the approach adopted by the U. S. National Highway Traffic Safety Administration.

Toronto's Traffic Data Centre and Safety Bureau maintains an electronic database containing information on all police-reported motor-vehicle collisions that occur in the city. It lists the date, time, and location of each incident, the age of the parties involved, the severity of any injuries sustained, as well as other basic information. Collision report numbers are also included for each entry, allowing electronic retrieval of the associated Motor-Vehicle Accident (MVA) reports, which contain many other important details. Most of the MVA reports contain a diagram depicting the configuration of the vehicles involved and a brief synopsis of the sequence of events, written by the investigating police officer. These were vital to the analysis of each incident.

After reviewing the available data, it was decided that collisions from the two most recent years for which records were available (1997 and 1998) should be analysed. Over 2,500 collisions involving cyclists were reported during this period, providing a robust sample size. Examining all the collisions that occurred over each complete twelve month period ensured that any seasonal variations in the prevalence of certain kinds of incidents would not affect the overall results. Some MVA reports were missing, while others were incomplete, so data for approximately 6% of the collisions is incomplete.

As each collision report was examined, the time of day, weather, and other basic facts were noted. The diagram and synopsis were carefully studied, in order to 'reconstruct' the incident. The collision was then assigned a "crash type" from the typology illustrated in table 2.1. The crash typology used here is based on one developed by the Federal Highway Administration (FHWA) for a 1996 collision analysis covering six U.S states (see Appendix A). That system consists of 38 different crash types, and is an updated version of a typology first developed by Cross and Fisher in 1977, for the U.S. National Highway Traffic Safety

Administration (see Hunter *et al*). The FHWA typology was used initially for this study, but after coding the data for one full year's collisions it became apparent that the classification system could be modified to better suit the local data. Our ability to compare findings with those of other studies that use the FHWA typology has not been significantly affected by the changes outlined below. (The relationships between the two typologies are detailed in Appendix A.)

The FHWA's category "Cyclist Strikes Stopped Vehicle" does not differentiate between situations where a cyclist simply runs into a parked car and those where a motorist suddenly opens a car-door, striking a passing cyclist. This type of incident is reported frequently in Toronto, so an additional category, "Motorist Opens Vehicle Door into Path of Cyclist," was created. On the other hand, some of the categories in the FHWA system pertain to collisions that are not frequently reported in Toronto ("Play Vehicle," "Wrong Way Motorist," and "Cyclist Right Turn," for example). These were collapsed into a miscellaneous category ("Other"). The resulting loss of detail was considered acceptable, since the small numbers of cases would have precluded any meaningful statistical analysis. Several other crash types in the FHWA system are distinguished by features that seem rather subjective: the categories "Motorist Overtaking — failed to detect cyclist," "Motorist Overtaking — cyclist path obstructed," and "Motorist Overtaking — misjudged passing space" all depend on statements from the driver. These were combined as a "Motorist Overtaking" category, reducing the total number of crash types to twenty-three, as listed in table 2.1. These categories are mutually exclusive and, along with the "Unknown" category for cases lacking complete information, account for all the cases.

Each crash type names the vehicular manoeuvres that most clearly describe the configuration of a particular kind of collision event. "Drive Out…" refers to the actions of the motorist, while "Ride Out…" refers to the actions of the cyclist. Where only one vehicle is mentioned in the crash type title, the other party can be assumed to have been proceeding "normally"(*i.e.,* straight ahead). For the cyclists, this could in some cases include riding along the sidewalk. Sidewalk cycling was therefore recorded as additional information necessary to fully describe the crash configuration. In some cases sidewalk cycling can also be considered a factor that may have contributed to the occurrence of the collision. This is because sight-lines at intersections (including driveways) are often restricted at the sidewalk. A cyclist crossing a roadway from the sidewalk, even at a moderate speed, can enter the motorist's field of view much more suddenly than a pedestrian would. Motorists scanning for pedestrians as they hastily negotiate an intersection may not expect to encounter cyclists in this part of the right-of-way. Cyclists on the sidewalk may not be able to see approaching motorists until the last moment, or may mistakenly assume that motorists have noticed them.

Environmental conditions, behavioural factors, and other external circumstances are often significant as well. Hence, as each collision was categorised, other factors that may have played a contributing role were also recorded. As the first few dozen reports were examined, a list of such factors that came up repeatedly was assembled (table 2.2).

Table 2.1

Collision Typology
1) Drive Out At Controlled Intersection
2) Drive Out From Lane or Driveway
3) Motorist Reversing
4) Motorist Right At Red Light
5) Motorist Right Turn (Not at Red Light)
6) Motorist Opens Vehicle Door in Cyclist's Path
7) Cyclist Strikes Stopped Vehicle
8) Motorist Left Turn – Facing Cyclist
9) Motorist Left Turn – In Front Of Cyclist
10) Motorist Overtaking
11) Ride Out At Mid-block
12) Ride Out At Controlled Intersection
13) Ride Out From Sidewalk
14) Ride Out From Lane or Driveway
15) Wrong Way Cyclist
16) Cyclist Left Turn In Front Of Traffic
17) Cyclist Left Turn – Facing Traffic
18) Cyclist Lost Control
19) Cyclist Caught in Intersection
20) Cyclist Overtaking
21) Drive Into/Out of On-Street Parking
22) Other (Not Classified)
23) Unknown

Table 2.2

Possible Contributing Factors
Cyclist riding on sidewalk or crosswalk
Darkness/poor visibility
Child cyclist (inexperience)
Sight lines obstructed
Motorist improper/unsafe lane change
Cyclist passing on right
Cyclist disobeying traffic control
Cyclist on wrong side of road
Motorist misjudged passing space
Motorist disobeying traffic control
Motorist discharging passenger in left lane
Cyclist path obstructed
Vehicular assault
Mechanical defect (bicycle)
Streetcar tracks
Cyclist impaired
Motorist impaired
Poor/wet road surface
Motorist failed to detect cyclist

Weather, light, and road surface conditions are examples of environmental factors that are routinely noted in the police reports. These data were recorded for analysis in every case, except when other factors eliminated the need — if the driver was charged with deliberately striking the cyclist, for example. Besides recording these variables as they appeared in the MVA reports (*e.g.,* “clear,” “rain,” “snow,” “freezing rain”), they were also treated as binary variables (*e.g.,* wet/dry), for the purpose of assessing their statistical role as possible contributing factors. Whenever possible, the distinctions referred to above, which define the different types of “Motorist Overtaking” crash types in the FHWA typology, were also noted as possible contributing factors.

Some behavioural factors, such as cycling on the sidewalk or on the wrong side of the road, were clearly apparent from the diagrams in the reports. Others, such as disobeying traffic control devices or failing to yield the right-of-way, were noted in code boxes, mentioned in the written synopses of the investigating officers, or could be inferred from charges laid. For each collision, as many as four of the possible contributing factors listed above were attributed. (In no case was it necessary or desirable to note more than four.) Recording these factors allowed for a detailed analysis of each type of crash, and compensated for any loss of information that might have resulted from combining some of the FWHA’s categories.

For each collision, the severity of any injuries sustained by the cyclist was coded for analysis. This allowed us to identify the types of collisions that tend to result in the most severe injuries (see section 3.2). Injuries are recorded in the collision reports according to the Ontario Ministry of Transportation scheme illustrated in table 2.3, except that the thirty-day limit on the “fatal” class is not strictly adhered to by Toronto Police.

Table 2.3: Categories of Injuries

Injuries	Description
None	Uninjured person.
Minimal	Person did not go to hospital when leaving scene of accident. Includes minor abrasions, bruises and complaints of pain.
Minor	Person went to hospital and was treated in the emergency room but not admitted.
Major	Person admitted to hospital. Includes person admitted for observation.
Fatal	Person died within 30 days, as a result of the collision.

2.2 Data Analysis

As the data was assembled and coded, it formed a matrix consisting of a row for each incident and a column for each data variable. (A complete list of all the variables collected for analysis appears in Appendix B.) After generating summary statistics for the entire data set, the data was sorted in various ways to create sub-sets. By comparing the statistical characteristics of each such sub-set with the characteristics of the whole sample, features peculiar to each sub-sample can be identified. This reveals relationships between some of the variables. For instance, if the data is sorted by the age of the cyclists involved, it is possible to identify the kinds of crashes that seem to pose distinct problems for each age group; when sorted according to geographic area, different problems stand out in different areas; if the data is sorted by crash type, it is possible to identify the age groups that are disproportionately involved in each type of collision. Sorting by crash type also highlights any statistically common factors that may have contributed to each particular type of collision, which can lead to suggestions for possible counter-measures. The main focus of our analysis took the latter route — comparing the statistical characteristics of the different crash types. The analysis was facilitated by the use of a series of histograms (bar charts) to compare the distribution of particular variables among different sub-sets.

To use the variable “cyclist’s age” as an example, if the ages of all the cyclists in any group are plotted in a histogram, it forms an “age profile.” The age profile of the group of cyclists involved in one type of collision can be plotted against the age profile for all the cyclists in the entire study sample, to reveal age groups that are “over-represented” (*i.e.,* more frequently involved in that particular type of crash). This process was repeated with the other key variables, including weather, light and road surface conditions, and all the “possible contributing factors.” The location of each collision was mapped, using Arcview GIS software, to facilitate the analysis of relationships between location, crash type, and the other variables. Statistics for each crash type were compared with statistics for the whole data set, using histograms in most cases. (Plotting the frequency distribution of a variable over one type of collision against its distribution over *all* collisions — as opposed to *all other* collisions — facilitates visual comparison of data for two or more crash types, since they are presented in relation to the same ‘base line.’) In this manner, the likelihood that each variable might have played a contributing role in each crash type was assessed.

While histograms provide a visual tool that is clear and easy to comprehend, more rigorous statistical tests are necessary to determine the significance of any apparent over-representations. For this purpose, cross-tabulations were performed for all the statistical statements made in this report, using SPSS software to apply Fisher's Exact Test. This was most important for crash types containing of a small number of cases, or for histograms exhibiting very small deviations from the base line. In cases where the sample size was large and a histogram clearly indicated a high degree of over- or under-representation, the statistical significance of the resulting statements was consistently found to be high.

2.3 Presentation of Analysis Results

The format used to present the analysis results (Chapter 4) loosely follows Carol Tan-Esse's treatment of the data from the 1996 FHWA study referred to above,[21] which includes a mix of histograms and tables. Histograms are particularly helpful for multi-category variables, like age and injury severity, and are used in this report to present detailed findings regarding these attributes, for every crash type. Histograms displaying environmental conditions have also been included, wherever an interesting result was found. For the "possible contributing factors," all of which are simple binary (yes/no) variables, any over-representations are highlighted simply by comparing a variable's frequency for a particular crash type with its average frequency over all crash types, in tabular format. Maps showing the locations of all the collisions of each type (for the two-year study period) are also included.

A useful feature of histograms is that each vertical bar can be made to represent a defined group, by adjusting the scale divisions (or "bin size") of the horizontal axis. Tan-Esse splits age data into seven groups: 0-9 (child), 10-14 (youth), 15-19 (teen), 20-24 (young adult), 25-44 (adult), 45-64 (mature adult), and 65 and over (elderly). Such formatting facilitates interpretation of the data, since the age distributions for each crash type can be easily visualised in terms of

[21] Tan, 1996

hypothetical "life stages." As an example, Figure 2.1 clearly reveals that the "adult" group is over-represented in a hypothetical "Crash Type X." The black bar representing cyclists age 25 to 44 involved in "Type X" collisions is much taller than its grey twin, which represents the same age group in collisions of *all* reported types. Of course, the definition of such age categories must have a basis in theory, or be supported by empirical evidence that different age groups consistently exhibit tangible differences in behaviour. Hypotheses about the cycling habits of different age groups can be tested by examining the degree of involvement of each age group in different types of collisions. With the help of histograms, many different crash types can be quickly and easily compared.

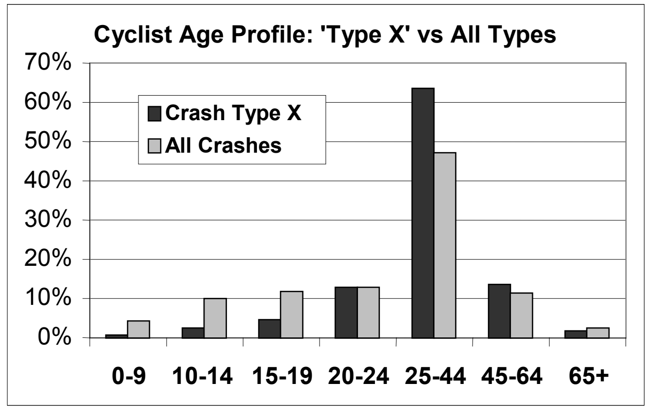

Figure 2.1

Another reason one might group distinct data, like the age of each cyclist, into a few relatively large intervals is to compensate for the random variations that can occur with small samples. On the other hand, when a *large* sample of distinct data is grouped into coarse categories, detail is inevitably lost. The FHWA study sample contained predominantly young riders, so it makes sense to split them into smaller age groups. However, almost half of the cyclists in the Toronto sample would fall into the FHWA's "adult" age group. This would hide any nuances that might arise *within* the "adult" category. For our sample, it is more informative to split the "adult" category into smaller sub-divisions. The large size of the bar representing the adult age group in Figure 2.1 also means that the differences between the other age groups are less striking, since they are compressed vertically. For these reasons, the Toronto sample was divided into uniform age categories, rather than hypothetical life stages. Because of the large size of the sample, many of the sub-samples (*e.g.* the seven most frequent collision types) contain sufficient cases to allow analysis using five-year intervals. Age data for the less frequent collision types, and other sub-groups containing fewer than 60 cases, is charted in ten-year age intervals.

2.4 Data Limitations and Reliability

The scope of this study was limited to an examination of collisions between motorists and cyclists that occurred within the city between January 1, 1997, and December 31, 1998, and were reported to the police. Estimates of the proportion of bicycle crashes that go unreported range as high as 90%.[22] While bicycle-only crashes and falls are rarely reported to police, studies of hospital records have also found that up to 60% of cyclists treated in hospital after a collision with a motor-vehicle did not report their collisions to police.[23] Collisions resulting in serious injuries are more likely to be reported than minor collisions involving little or no property damage or injuries. Thus, police records must certainly under-represent the frequency of car/bike collisions, while over-representing their average severity.

The MVA reports, the primary source of our information, are of two types. Most are prepared by police officers at the scene of the crash. Basic information is entered in a standard form (see Appendix B), which also has space for a diagram and written synopsis, where the officer provides his or her interpretation of the incident. A few (less than 5%) were prepared at Collision Reporting Centres, and include the statements of one or both of the parties involved. In both situations, there is potential for error, bias and/or misinformation. The study did reveal some shortcomings in the collision reporting forms. (Data on the use of bicycle lights, reflectors, helmets and other safety equipment, for example, is not collected consistently.) It also highlighted a few areas where police officers' awareness and appreciation of bicycle traffic safety issues might be improved.

In cases of inconsistent or conflicting information, the primary investigator and the research supervisors were usually able to arrive at a plausible re-construction of the incident together. Indeed, it is important to remember that in *every* case, the crash-typing process involves a degree of re-construction. While the MVA reports contain a great deal of information, they are geared primarily for dealing with the sort of issues that arise following collisions between motor-vehicles — legal and insurance issues, and such. Thus, many details that might interest the

[22] Aultman-Hall and Kaltenecker, 1999

[23] Stutts and Hunter, 1999

researcher are simply not available. The cyclist's perspective on what happened is particularly absent from this kind of data.

The collision typology was tested for reliability by having four separate investigators categorise a representative sample of the collisions. Even with minimal training, the investigators agreed on the choice of collision type in approximately 80% of cases. The size of the study sample (over 2,500 cases) was sufficient to allow for confident statements regarding the characteristics of the most frequent types of crash, and the relationships between many general factors. When presenting findings for collision types where the number of cases is small, low statistical confidence is noted, even if the patterns that emerged seemed consistent with 'common sense.' A larger database would provide a higher level of confidence at all levels, but especially for the investigation of interesting but small sub-samples.