

Making Sense of Injury Data

Hassan Soubhi, Ph.D.

British Columbia Injury Prevention
Conference, Sidney, May 25, 2001

Understanding and interpreting data is the process by which we make sense of data. Such a process involves various ways of looking at the data. In this workshop, we will examine one of the most prevalent ways of looking at injury data, using examples from injury mortality and hospitalization data in British Columbia.

Data? What are they?

Data can be defined in two ways¹: First, as factual information that can be used as a basis for reasoning or discussion. Second, as any information that can be collected by a sensing device (e.g., tape recorder, questionnaire). Such information includes both useful and irrelevant or redundant information and must be processed to be meaningful. This second definition is most useful to our purpose, since it highlights the need to process the information contained in the data in order to understand it and interpret it.

What is Information?

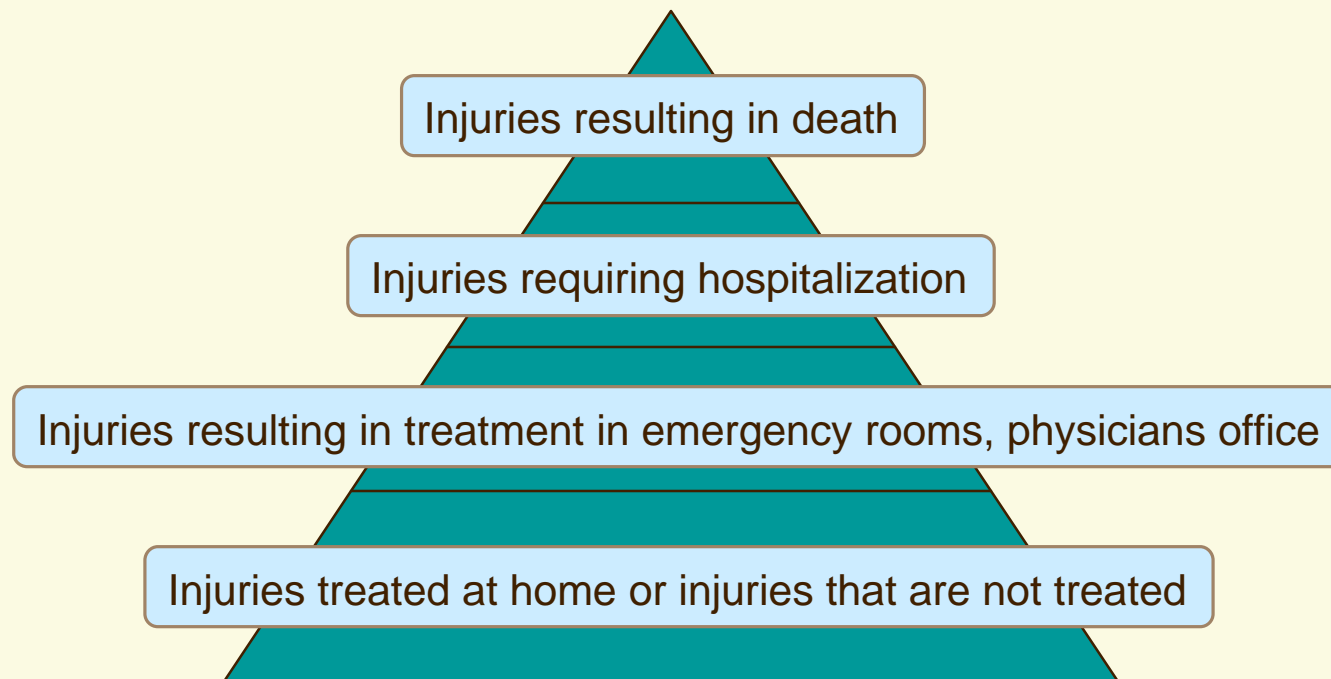
Information can be defined as that which surprises our senses and to which we attach meaning. Based on the previous comments on the definitions of data, information can be seen as the complement of noise or irrelevant information in the data. In other words:

$\text{Data} = \text{Information} + \text{Noise}$

Since information is what we are looking for in the data, the first step in any attempt to understand or interpret data is therefore to get rid of noise (irrelevant information, coding errors) in the data.

What Data Should Be Collected?

Injury Pyramid



The Injury Pyramid is a visual representation of the levels of injury severity and the various potential sources of injury data. Sources and types of data include for example, injury mortality from BC Vital Statistics, hospitalizations from the BC Ministry of Health, and Emergency Department data from CHIRPP.

The level of detail about the injury event varies depending on the sources of data. Ideally, injury data should include answers to all of the following questions: What, how, when, where, who, why.

What Data Should Be Collected?

- ✓ The triad host/agent/environment provides a useful framework for collecting injury data. This framework includes factors describing the injured person, the agent that caused the injury, and the environment where the injury occurred.

What Data Should Be Collected?

- ✓ Characteristics of Host
 - age, gender, community of residence
- ✓ Causal Agent
 - ICD-9 or ICD-10 E-codes / N-codes. These codes provide an indication of the type of energy exchanged during the injury (e.g., mechanical, thermal, chemical, etc.) and the nature of injury
- ✓ Environment
 - physical and social environment

Example of a Minimum Data Set

- ✓ Date of Birth
- ✓ Date of Visit
- ✓ Gender
- ✓ Postal Code
- ✓ Main Problem (N-code)
- ✓ Other Problem (N-code)
- ✓ External Cause (E-code)
- ✓ Place of Occurrence of Injury
- ✓ Narrative Description of the Injury Event
- ✓ Activity When Injured
- ✓ Visit Disposition

Of Rates and Proportions

Data that are collected based on the preceding framework contains information at the individual level. Individual level data are useful for clinical care to a specific individual.

Injury control is geared towards groups of people and the best way to describe a group is by using rates and proportions².

- ✓ An injury rate expresses the number of injuries during a stated period divided by population size. It is always recommended to divide the number of injuries by the population that is at risk of having the injury.
- ✓ For example, Injury Death Rate = Injury Deaths during a year/Mid-year population. If during a given year in a given area, 20 individuals aged 25-34 years died because of motor vehicle collisions, and we know that by mid-year there were 5,256 individuals in that age group, then the death rate is: $20/5,256 = 0.0038$. Usually a rate is expressed per 1,000 (or 100,000), so our death rate becomes: $0.0038 * 1,000 = 3.8$ per 1,000 population.

- ✓ A proportion can also be used to represent the relative importance of a given injury. For example, 2 injured children out of 10 every month in a given school is more alarming than 2 out of 400.
- ✓ Finally, we should note that a proportion does not have any unit, and its values can range from 0 to 1 (or 100%). A rate on the other hand is expressed per unit population and can range from 0 to infinity.

Making Sense of Data

We can now define making sense of injury data as the process by which we identify relevant and useful information about injuries. Such a process includes using the framework host/agent/environment to collect data from a specific source for a specific level of injury severity (e.g. injury pyramid). Once we have the relevant information from the relevant data, we can analyze it to identify:

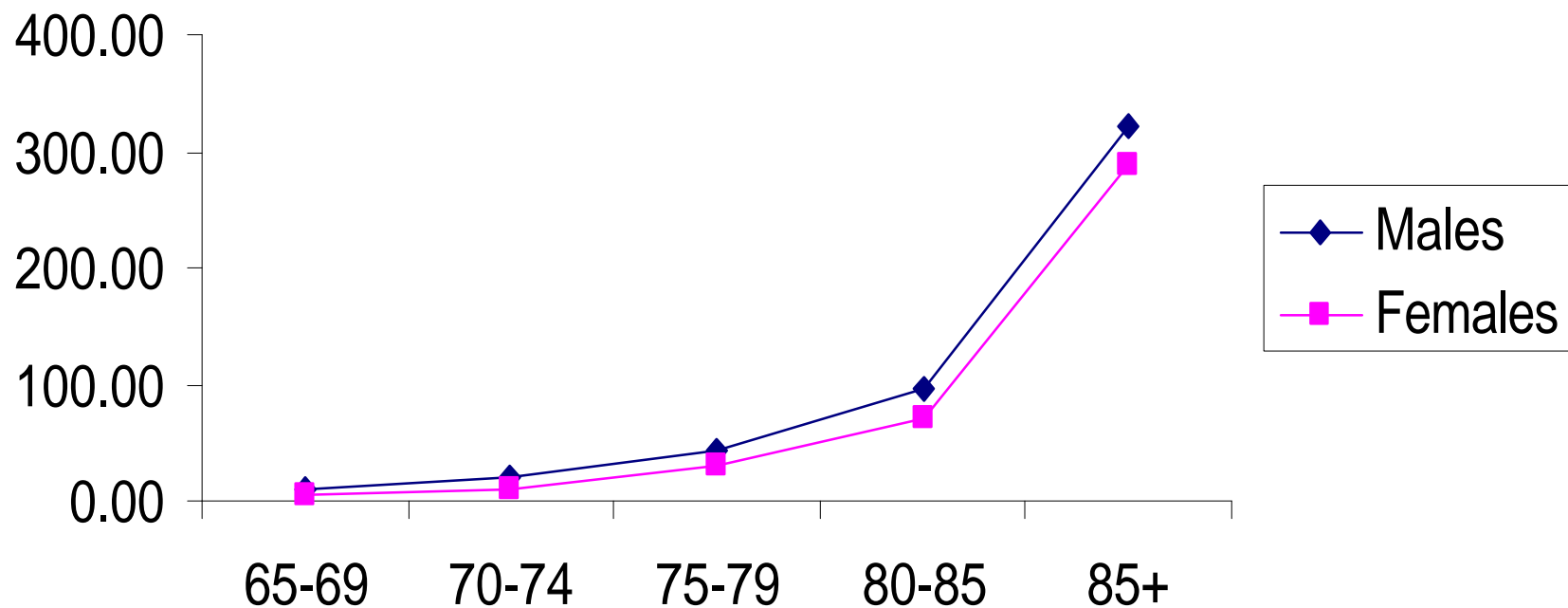
- * Groups of people that share specific characteristics such as age, gender, occupation, type of exposure (e.g. activity when injured)
- * Time trends using date of visit for example
- * Characteristics of places where injuries occur: e.g. regional variations

The next slides present some examples using mortality and hospitalization due to falls among seniors and elderly in BC.

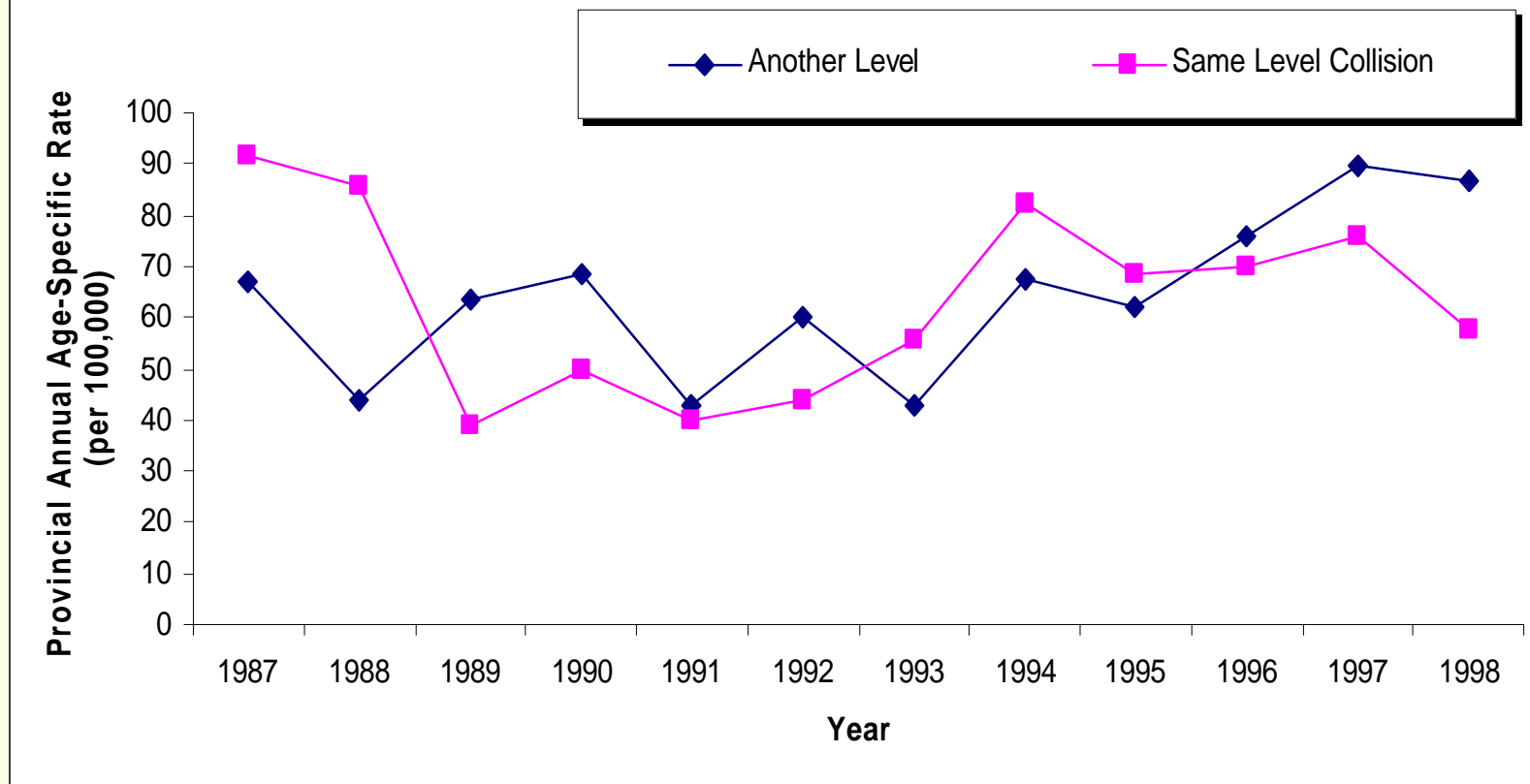
Average Annual Age-Specific Mortality Rates per 100,000
and Number of Deaths due to Falls, B.C., 1987-1998

Age group	Male		Female	
	Total # of Deaths 1987-1998	Rate	Total # of Deaths 1987-1998	Rate
65-69	96	10.49	57	5.61
70-74	148	20.44	92	10.17
75-79	214	41.81	208	29.31
80-85	290	95.29	335	70.16
85+	608	320.29	1135	288.35
65+	1356	51.26	1827	52.18

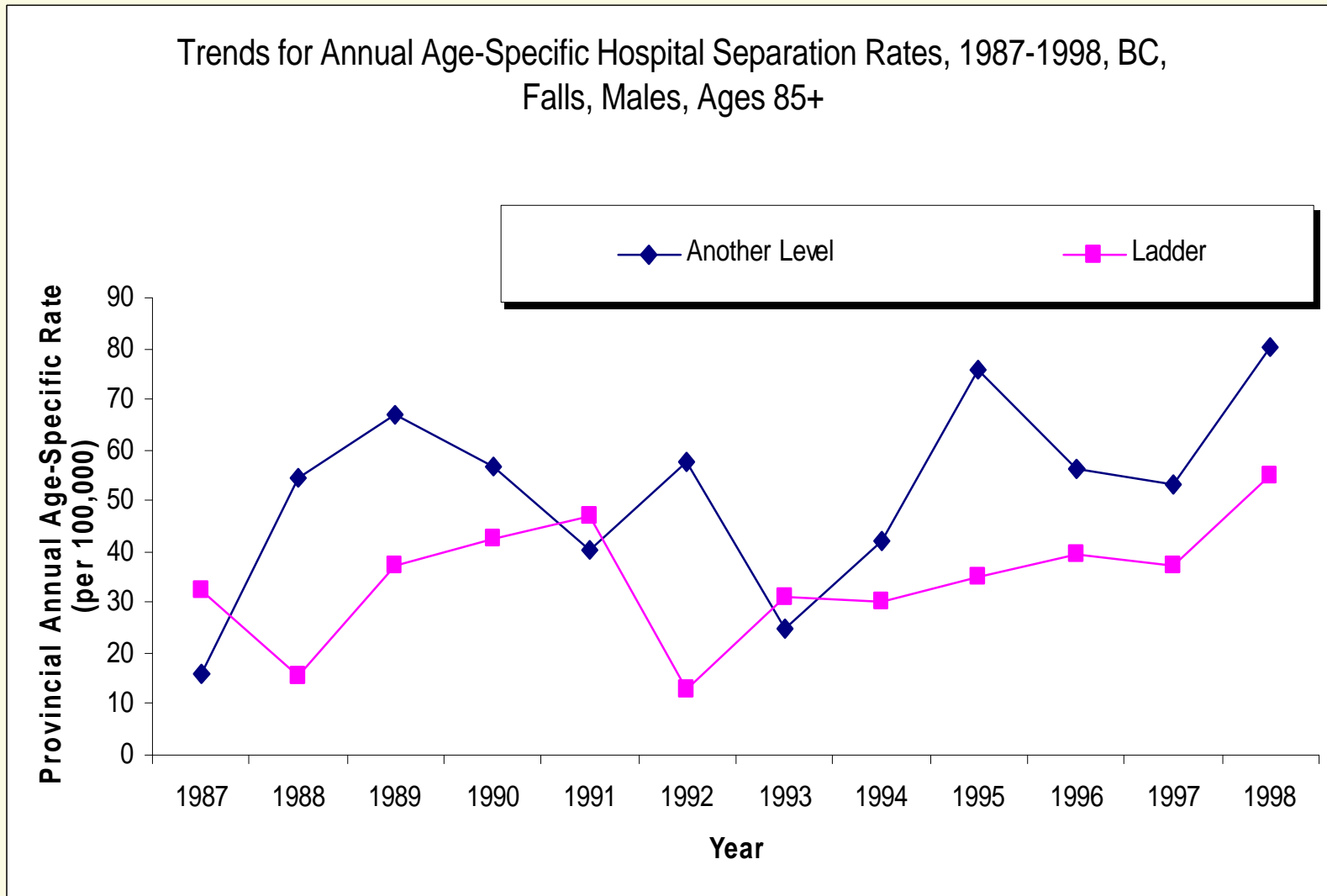
Age-Specific Mortality Rates per 100,000, Falls, BC, 1987-1998



Trends for Annual Age-Specific Hospital Separation Rates, 1987-1998, BC,
Falls, Females, Ages 85+



Trends for Annual Age-Specific Hospital Separation Rates, 1987-1998, BC,
Falls, Males, Ages 85+



Comparisons: Differences and Ratios

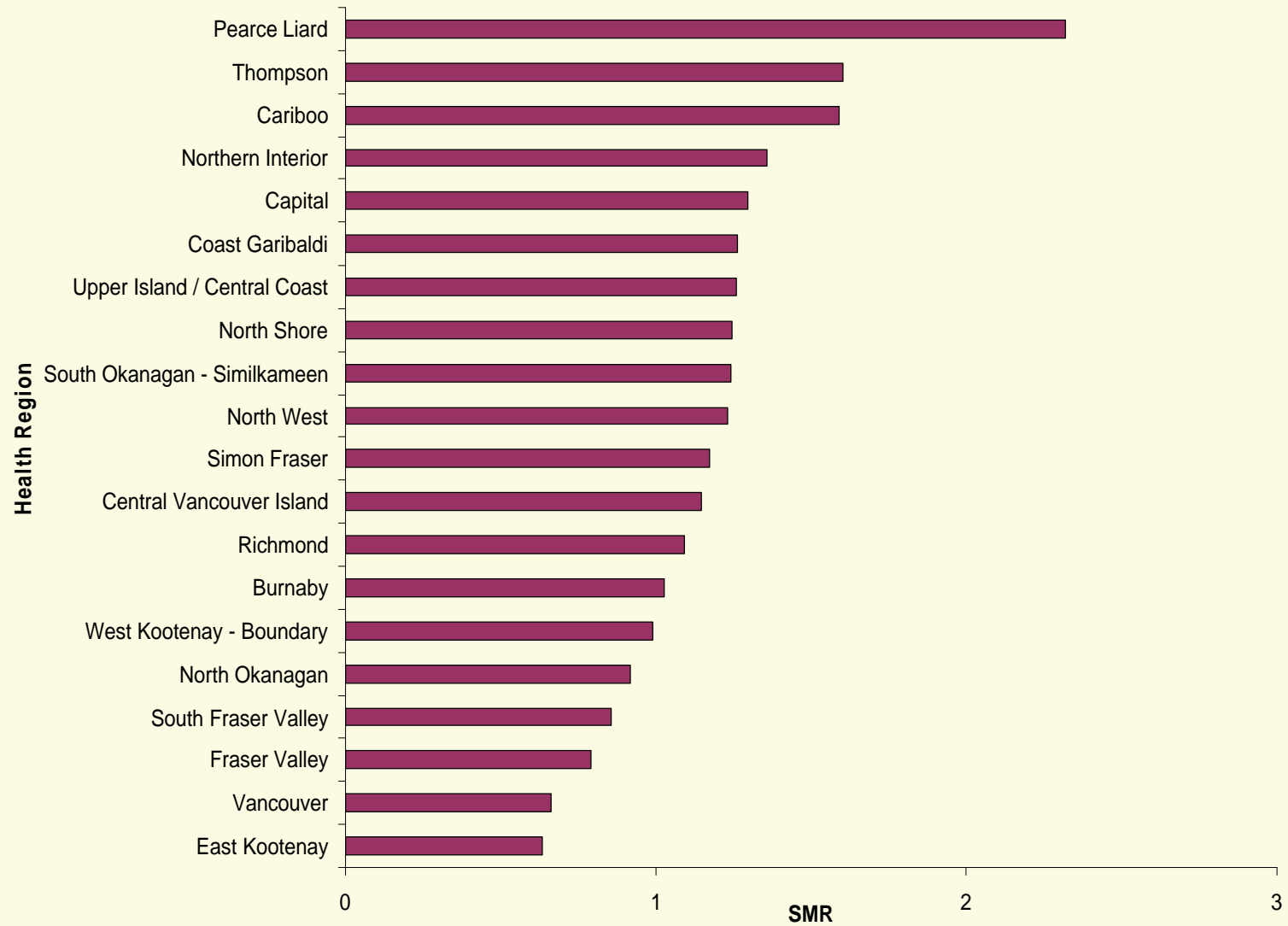
We usually need to make comparisons between different groups to identify differences in the rates of injury and the characteristics of the groups that could explain those differences. For example, we are often interested in detecting differences between boys and girls, or between injuries in the home and out of the home.

There are at least two ways to conduct such comparisons:

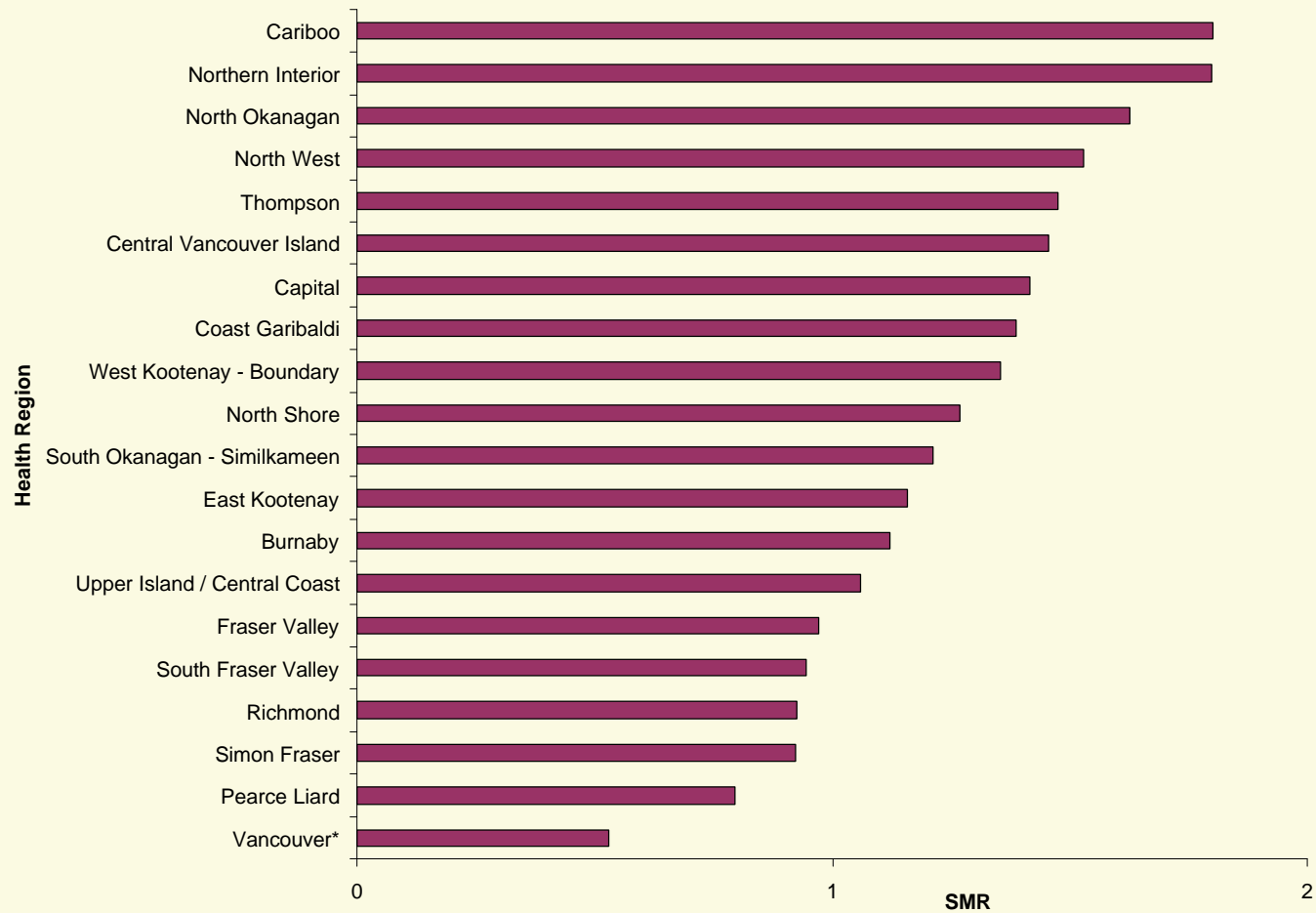
1. Calculate the difference in injury rates between two groups, e.g. boys versus girls
2. Divide the injury rates of the two groups to obtain a rate ratio.

The next slides show an example of regional comparisons separately for males and females using Standardized Mortality Ratios (SMR) to take into account differences in the age structure across health regions. All the comparisons are made relative to the province (SMR = 1).

Standardized Mortality Ratio, 1987-1998, BC,
by Health Region, Males, Falls, Ages 65+



Standardized Mortality Ratio, 1987-1998, BC,
by Health Region, Females, Falls, Ages 65+



Making sense of injury data also involves separating what is real from what could be due to chance or bias.

* Chance events are random events. There are a number of statistical tools we can use to identify the results that are statistically significant (i.e. not due to chance)^{3,4}.

* Bias represent a systematic error in the data that is the cause of erroneous or inaccurate results. For example, surveys using self-reported information can be subject to recall bias where the individuals reporting about injuries may under or overstate some injury events for various reasons including lapses of memory. The result can be an under or overestimation of the true injury rates.

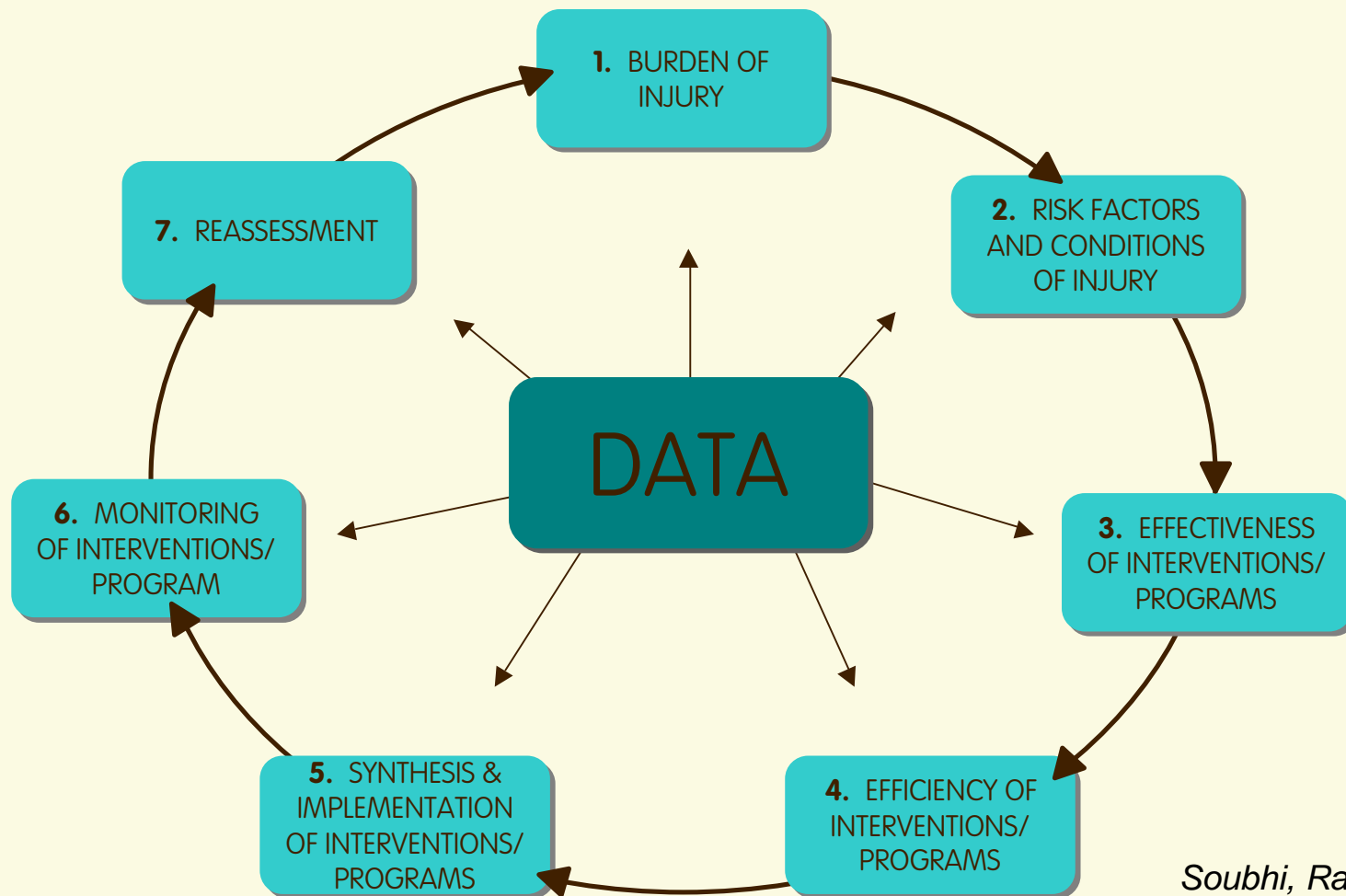
In summary, making sense of data requires answers to the following questions:

- * Where do the data come from?
- * How accurate are they?
- * Which part of the pyramid is reported?
- * Which group or population is reported?
- * Does the data contain information about the host, the agent and the environment where the injuries occurred?
- * How biased are the data?
- * How to analyze the data (comparisons)?
- * How confident can one be in the results? (Issues related to bias and statistical significance)

Ultimately, what do we want data for?

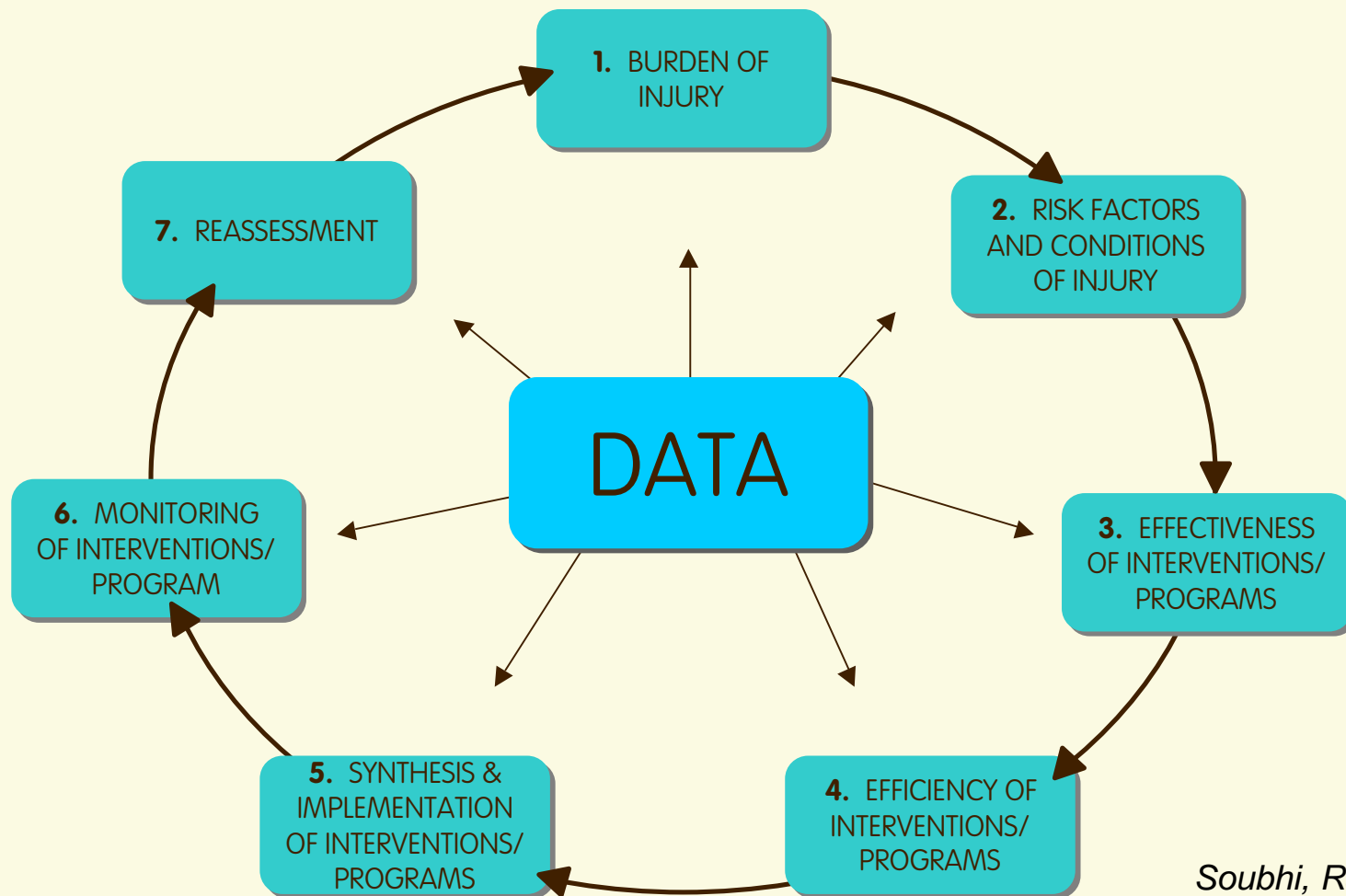
Data are part of a broader cycle of assessment and evaluation. This cycle is represented in the following slides as the Injury Prevention and Evaluation Cycle (IPEC) ⁵. Data keep records of information about each of the steps and help establish links between them. Data can then be represented as the hub of the cycle, since data are essential for each of the steps represented in the cycle.

Injury Prevention Evaluation Cycle (IPEC)



Soubhi, Raina et al,
1999

Data: The Hub of the Wheel



Soubhi, Raina et al,
1999

References and Suggested Reading:

1. Last, J.M. (Ed.). (1995). *A dictionary of epidemiology*. New York: Oxford University Press.
2. Friedman, G.D. (1987). *Primer of epidemiology*. Third edition. New York: McGraw-Hill.
3. Norman, G.R., & Streiner, D.L. (1986). *Statistics*. Toronto: BC Decker, Inc.
4. Hirsh, R.P., & Riegelman, R.K. (1992). *Statistical first aid. Interpretation of health research data*. Boston: Blackwell Scientific Publications.
5. Soubhi H, Raina P, Chong M, Turcotte K, Babul S, Olsen L, & Scanlan A, (1999). *Unintentional injuries in British Columbia: Trends and patterns among children and youth 1987-1996*. Vancouver, BC: BC Injury Research and Prevention Unit.