CANADIAN HANDBOOK ON HEALTH IMPACT ASSESSMENT

Volume 3

Roles for the Health Practitioner

DRAFT

DECEMBER 1999

Biostatistical Concepts and Methods

Chapter G Contents

References

Introduction

Biostatistics are used in Health Impact Assessment (HIA) and Environmental Impact Assessments (EIA) to quantitatively describe present health, social and economic status and the predicted impacts of the project being assessed. But what are statistics and why and when do we use them?

A statistic is a value which summarizes some quantifiable aspect of data. Often the data we observe are a subset or *sample* of a much larger *population* of data. Samples are simpler and less costly to study than populations. However, the objective of the study is usually to make statements about the population. This can be accomplished with samples as long as the sample is *representative* of that population. The key is to understand how the sample relates to the population. There are many, often infinite ways of selecting a sample. Thus, any quantified statements about the sample become *probabilistic* statements about the population. Selecting the sample, analysing the data, and drawing inferences about the population is a complex activity which is the subject of the discipline known as *statistics*.

A *statistic* is a value which summaries some quantitative attribute of the data. For example, the mean is the average value of the data.

Definition:
Statistics is a discipline dealing with methods and rules of obtaining data, analysing and summarizing it, and drawing inferences from data samples by the use of probability theory. It is governed by the *laws of probability*, and involves numbers and randomness but requires *logic* and *problemsolving*.

Useful properties of statistics are that they are representative of the population and that they are reproducible to a certain degree. Representativeness is accomplished by randomly selecting a sample from

The entire set of objects that are of interest is called a *population.*
A portion of a population is called a *sample.*

the population. For example, in cleaning up a hazardous waste site, the concentration of contaminants throughout the site is valuable information. However, to measure the concentrations at every possible location is not practical so a sample of locations is selected. This sample, if carefully chosen, can provide statistics that are representative of the overall contamination of the site. Some kind of randomization in the selection of the sample, whether it is unrestricted or restricted by certain constraints, is the basis of ensuring representativeness. Statements can even be made about the confidence one has that the sample statistic represents the theoretical population statistic using confidence intervals.

Randomization is key to selecting a sample which is representative of a population.

Measurements on a population of items under study vary; that is, they are not a predictable constant. Thus, there is an underlying *distribution* for the measurements of the population. A *histogram* of the selected sample of measurements provides a reasonable picture of the population distribution. One of the more common distributions is the Normal distribution which is symmetric and bell-shaped.

Distributions are summarized using key statistics. For example the normal distribution can be completely summarized by the mean and variance; that is, the population distribution is estimated by estimating the mean and variance of the sample of measurements. Summarizing population distributions by a type of distribution and the statistics that specify the distribution allows us to make comparisons among populations.

Selecting a representative sample is not always straightforward. Constraints may dictate restrictions on the way the data is collected. Special sampling designs are developed in order to make the most effective inferences in the most efficient way. In the laboratory setting factorial, randomized block, and randomized incomplete block designs are examples of standard protocols which are selected depending on the objectives and

constraints.

The application of statistical methods usually requires that assumptions be made about the data. For example, to use the t-statistic for testing the hypothesis that the mean of a population equals zero is valid only if the data come from the same normal distribution and are independent of each other. If the methods are applied without consideration of the underlying assumptions upon which they are based, erroneous and misleading conclusions could result.

As the complexity of the sampling protocol and the hypotheses to be tested increases, the complexity of the methods for analysing the data increases as well. Statisticians are specifically trained through university programs to learn the proper application of complex statistical methodology highlighting the fact that statistics is an important scientific field of study. Statisticians with Masters degrees and Ph.D.s are trained to develop new and adapt old methods when required.

Some training in statistics for non-statisticians involved in HIA is recommended in order to understand the concepts, though it may not be necessary for a researcher or evaluator to know how to apply the methods themselves. Sometimes it is only necessary to know when a statistician should be consulted. Training can range from short, introductory courses to graduate level degree courses. Many disciplines such as toxicology, psychology and economics require at least one introductory statistics course in their degree program.

Statistical analysis has become much more accessible to non-statisticians with the advent and continual expansion of computing resources. Easy to learn, use and inexpensive, programs are now available on personal computers which are capable of relatively sophisticated analyses. Some are easy point-and-click programs while others are highly

advanced programming languages such as the Statistical Analysis System (SAS, 1990) and S-Plus (MathSoft, 1990). We reemphasize however, that care must be exercised in using these programs in order to ensure the integrity of the results.

Biostatistics in Health Impact Assessment

Opportunities for using statistics arise in all aspects of HIAs. Common applications include: evaluation of exposure; evaluation of fate of contaminants; evaluation of risk; projected costs to mitigate impact; evaluation of baseline exposure versus predicted future exposure; estimation of magnitude/impact; estimation of population impacted; quantification of health indicators; and levels of toxic chemicals in human tissues (see vol. 1).

There are several advantages to using reliable statistics in HIA:

- Using appropriate statistical methodology enhances the credibility of the results;
- Sampled data can be extrapolated to make inferences about a population and a level of confidence determined for the extrapolation;
- Statistics enables the assessment of variability and highlights uncertainties in the data;
- "Averages" can be estimated in light of the shape and location of the sampling distribution;
- Results can be interpreted in light of estimated variability.

There are also limitations to the utility of statistics.

- Even the most careful use of statistics does not mean that all the

A **Type I** error is said to have been made if we infer that there is a real difference between the two samples, while in fact the observed difference is due to chance only.

[Kennedy and Neville, 1974.]

answers will be provided. Statistics are probabilistic in nature and therefore carries with it a probability of making errors (TYPE I and Type II).

- The statistics are only as good as the data. Potential sources of error include: the data are not collected to address the specific objectives of the study; the measurement technology is unreliable or highly variable; other confounding factors lead to misinterpretation of the results.
- Statistical analysis is highly dependent on assumptions. If one or more of the assumptions are invalid, then the conclusions could be erroneous.

A **Type II** error is said to occur if we conclude that there is no real difference between two samples, while the difference does in fact exist.

[Kennedy and

Statistical Analysis

Describing Data

Before analysing data, the structure of the data and how it was collected is determined. There are two key questions:

(1) What was the experimental or survey design?
(2) What kind of data was collected?

What is the experimental or survey design?

A well-designed experiment has many advantages. It can facilitate the use of valid and efficient statistical procedures. Results are often easier to analyse and interpret and more comparable with those of other studies.

Statistical principles of experimental design include randomization, efficiency, blocking, and structure. Randomization provides statistical independence and validity. It insures that nothing distinguishes the treatment groups other than the treatments themselves. For example, in animal toxicology experiments, to conduct proper randomization, treatments must be assigned randomly to animals and animals randomly assigned to cages. Blocking refers to blocks of experimental units. For example, if some experiments were done on one day and some on a different day, then day is a block effect. Structure refers to how the experimental units were layed out such as number of factors or time sequences for experimental units.

What kind of data was collected?

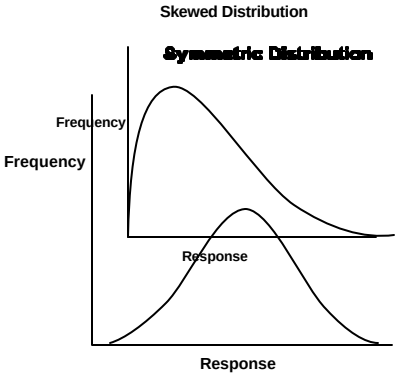
Many end points may have been collected during the course of the experiment. There are three types of outcomes:

- continuous
- proportions
- count

Examples of continuous endpoints are body weights, feed consumption, clinical chemistry, haematology, urinalysis results, and organ weights. Proportions arise when the number of animals with tumours in a group of animals is of interest. Examples of count endpoints are white blood cell counts and number of mutations in a mutation assay.

It would be convenient if there was a one-size-fits all statistical analysis, but this, unfortunately, is not the case. Procedures that may appear to fit many situations may not be the most efficient tests and lack power to detect effects. Thus, great care must be exercised in choosing the right analysis. The experimental design and the type of data determines the appropriate statistical analysis.

Estimation



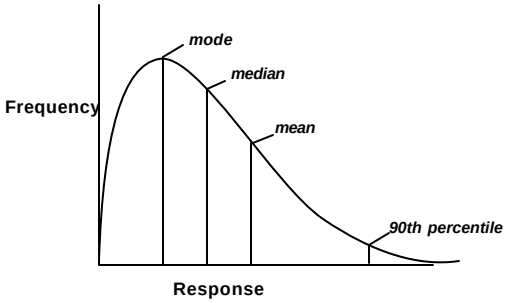
A statistic is a function of the data meant to summarize a particular aspect of the distribution of the data which can be symmetric, skewed, bimodal or some other shape. For a normal distribution, if the mean and the variance are known then the entire distribution can be reconstructed. The true mean and variance may be unknown but can be estimated using the statistics, the sample mean and variance which are calculated using the data. The mean is a measure of central location for the normal distribution. Other measures of central location are the median, mode and geometric mean. Other statistics which help describe a distribution are percentiles (or quantiles) and a variety of measures of variability such as the variance and the range of data.

Definitions of commonly used summary statistics	
Sample Statistic	Definition
Mean	**sum of all data divided by the number of data points (n)**
Median	**data point where 50% of the data have smaller values**
Geometric mean	**product of all data to the power 1/n**
Variance	**sum of squared deviations of each data point from the mean, all divided by number of data points minus one**
Standard deviation	**square root of the variance**
Standard error of the mean	**square root of the variance divided by the number of data points**
nth Percentile	**The data point where n% of the data have smaller values**

n% Confidence interval for the mean	**The confidence interval covers the true mean, n% of the time.**

An issue that often arises in exposure assessment is determining a meaningful summary measure of exposure if the data come from a skewed distribution. If there are a small number of very large values, a frequent occurrence in exposure data, then the mean may be greatly influenced by these points and thus may not accurately summarize the central tendency of the distribution of data. The median and the geometric mean are more insensitive to these large values. A decision about what measure to use will be based on the objectives of the assessment.

Descriptive Measures of a Skewed Distribution



Hypothesis Testing

Making comparisons among populations is done using a rigorous statistical decision theory approach. In the classical (Neyman-Pearson) approach, a *null hypothesis* is constructed then tested against an *alternative hypothesis* using a *test statistic*. The null hypothesis is always stated as no difference or no effect. The alternative hypothesis is stated as all possible alternatives.

Example: The null hypothesis for testing that the means of two populations, F_1 and F_2, are equal is written as

Ho: $F_1 = F_2$

which is tested against the alternative

Ho: $F_1 \ldots F_2$.

This is a two-sided test. For a one-sided test, the … will be replaced with < or > as appropriate.

Detecting Statistically Significant Effects

Comparisons among groups

Methods for establishing statistical significance of effects depend on the type of data.

Continuous Data

Statistical tests depend on a number of assumptions that are made about the data. It is important to ensure the validity of these assumptions in order to ensure the validity of the statistical test. Suppose that 10 animals were exposed to a substance and there were 4 such exposure levels. The objective was to determine if there was a detectable change in weight. Here weight is continuous. Analysis of variance (ANOVA) is usually used to determine if there are statistically significant differences between the means of the four dose groups. However, there are three assumptions that must be made about the data to ensure the validity of the analysis of variance procedure.

(1) the data are independent from one another,
(2) variance of the measurements are equal, and
(3) weights are normally distributed.

Independence can usually be ensured if strict randomization of the allocation of animals to treatments is carried out. Normality would be a characteristic of the data. This can be

validated in a number of ways. If the data come from the same normal distribution then a histogram of the data will have a symmetric, bell shaped curve. One common misconception is that the raw data from an experiment should have this shape if they are from a normal distribution. This is not necessarily true if there are differences among the means of the dose groups. In this case, a histogram may appear bumpy indicating more than one mode in the data. To check out normality, the group mean can be subtracted from the all the data within that group. A histogram of the resulting *residuals* should then appear bell-shaped with only one mode if the data come from normal distributions.

Normality of Residuals
C **Histogram should be bell-shaped, not skewed**
C **Plot of residuals vs predicted values should have equal variability across predicted values**
C **Test of normality should not be significant**

Sometimes, after adjusting for the means, the residuals appear to be skewed with long tapering tails. This is usually the case for concentration data. Since the ANOVA procedure requires normality, the data needs to be transformed to normality before the test can be done. Applying a log transformation to concentration data often normalizes the data. Other types of data may require alternative transformations. Once normality is achieved, statistical tests are then conducted using the transformed data. If the sample sizes are large, then it is less important to achieve normality since the distribution of the means tends to become increasing normal as the sample size increases.

Provided the assumptions discussed above are applicable, the Student's t-test is used to test the hypothesis that two groups means are equal. To test the equivalence of the means

of more than two groups, ANOVA is used.

Significance tests for equality of means with normal data		
Test	Description	Assumptions
T-test	**Equality of means of two independent groups of data**	C **normality** C **independence of all data** C **homogeneous variance**
Paired t-test	**Equality of means of paired groups of data. Equivalent to a one-sample t-test on the differences between the pairs.**	C **normality** C **independence of all pairs of data** C **homogeneous variance**
ANOVA	**Equality of means of more than two independent groups of data**	C **normality** C **independence of all data** C **homogeneous variance**

The ANOVA procedure will establish that at least two means are statistically different from each other, but it will not indicate which means are different. To determine which means are different, multiple comparisons between the means are conducted. This is not as simple as performing pairwise t-tests between the control group and each dose group since the significance levels for the hypothesis tests are no longer valid. Instead, a multiple comparison method can be used. The appropriate method for the comparison of the control to all others individually is Dunnet's test. Duncan's Multiple Range test, another multiple comparison method is appropriate for pairwise comparisons among all possible pairs.

Multiple Comparisons		
Dunnet's	**Compare many treatment means to a control**	C **normality** C **independence** C **equal variances**
Duncan's	**Compare all possible pairs of treatments**	C **normality** C **independence** C **equal variances** C **equal group sizes**

Categorical Data

Categorical data arise when the response can be categorized into one of a small number of categories. For example, an animal in a carcinogenicity study is classified as either having a tumour or not having a tumour. The number in each category can then be counted. An example of such data is given below.

Binomial Data		
Group	Number of Animals with Tumors	Total Number of Animals
Control	**1**	**100**
Test	**4**	**50**

Note that the number of animals without tumours is the total number of animals minus the number of animals with tumours. Fisher's exact test is used to compare the proportion of animals with tumours in the control group to the proportions in the test group.

There are studies in which outcomes fall into more than two categories such as eye irritation and dermal irritation tests in which a score is given to the degree of irritation. For such set-ups, more sophisticated analysis is required. The SAS package contains the CATMOD procedure which will handle a wide range of analyses of categorical data (SAS, 1990). Shoukri and Edge (1996) have a good discussion of some of these techniques.

Count Data

Count data arise when the number of outcomes is of interest and there is no maximum number of outcomes. The number of mutations in a mutation assay is an example. Other examples of count data are white blood cell differential data and nucleated red blood cell counts. Counts have a Poisson distribution. A simple approach for analysing count data is to compute the square root of each count then use normal test techniques as discussed in Section 2.2. This works quite well when the counts are relatively large. If the counts are small, more sophisticated methods using the Poisson distribution allowing for over-dispersion due to experimental error would be more appropriate.

Non-parametric tests

The normality based test for continuous data is called a *parametric test* since the data is assumed to arise from a known distribution, namely the normal distribution. Sometimes, however, the data do not arise from a known distribution. For example, with continuous data, there may not appear to be any suitable normalizing transformation. The only assumption that may be reasonable is that the shape of the distribution for each exposure group is similar, but may be shifted. In this case, nonparametric tests can be used to test for significant shifts in location.

The Wilcoxan rank-sum and the Mann-Whitney U tests are used to test if two groups have the same location parameters. The Kruskal-Wallis test is used to test for differences among more than two groups. These tests are all based on the ranks of the data rather than on the data themselves.

Regression analysis

When the response is of the continuous type, for example, weight or concentration, and of interest is the relationship between the response and a second continuous variable, for example, height, then regression analysis is used to determine if there is a significant relationship and the form of the relationship. A simple linear regression where there is only one input variable is expressed as:

$$y = a + ß\,x + e$$

where y is the response, x is the input variable, a is the intercept parameter, ß is the slope parameter and e represents an error term which is often assumed to have a normal distribution with 0 mean and some unknown variance. The slope parameter can be increasing, decreasing or remain level. The intercept parameter is the average value of y when x is zero.

The simple linear regression situation can be generalized to include more input parameters, known as multiple regression. It could also include categorical input parameters. A mixture of an analysis of variance and regression set-up is called a general linear model. One of the requirements for the input variables is that they are independent of each other. If they are not, then multi-colinearity occurs and the analysis may not produce well- defined results.

The above assumes that the response is linearly related to the input variables. If this is not the case, then non-linear regression must be applied. For example:

$$y = a \exp(-ßx)e$$

is a common non-linear expression. Non-linear regression analysis would normally be performed using a statistics analysis package. In this case however, linear regression could be used if the model was transformed by taking the logarithm of each side of the equation so that:

$$\log y = \log a - ß\, x + \log e.$$

The analysis is performed on the new log y response variable, an easier approach than non-linear regression which is a technique requiring advanced knowledge of statistical procedures. The difficulty in analysing transformed data is in the interpretation of estimated models.

Once again, assumptions made about the data must be valid before performing the analysis. Linearity, normality, independence and homoscedasticity can be checked out on the residuals of the analysis. Neter et al (1990) have a comprehensive discussion of many aspects of general linear models.

Regression analysis can also be performed on categorical response data. The technique called logistic regression is explained more fully in Shoukri and Edge (1996). SAS contains a logistic analysis procedure (SAS, 1990).

Multivariate analysis

Multivariate analysis is applicable when there are several responses to analysis and the objectives are not only to determine relationships between input variables and responses but those amongst the response variables. Common multivariate techniques include principle component analysis, cluster analysis, factor analysis, and multi-dimensional scaling. A complete discussion can be found in Johnson and Wichern (1992). Computer programs such as SAS are particularly useful (SAS, 1990).

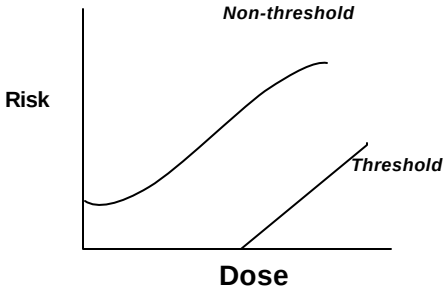
Special Topics

Estimating potency

Threshold toxicants

Guidelines and priorities are established on the basis of potency estimates of toxicants. There are two types of mechanisms for toxic effects that determine the method for estimating potency. Some effects are believe to have a threshold exposure below which the effect will not appear. Others such as genotoxic carcinogens, are presumed to have no such threshold.

Dose - Response Shapes



The figure to the right shows theoretical dose-response relationships for threshold and non-threshold toxicants.

The concept of a threshold centers around the

assumption that the dose level determines whether or not a pathological effect will occur upon exposure. Ideally, for the purposes of risk management, one would want to estimate this threshold. However, these thresholds may vary among individuals. In the uncertainty factor approach to risk assessment for threshold toxicants, a no observed effect level (NOEL) is selected based on a dose-response experiment in animals. Due to uncertainties of interspecies extrapolations from test animals to humans, and because of the individual variability in human sensitivity to the effects of toxic agents, safety or uncertainty factors (UF) are introduced. Thus, for example, an acceptable daily intake (ADI) may be computed by dividing the NOEL by the uncertainty factors UF, that is:

$$ADI = NOEL / UF$$

Uncertainty factors range from 1 to 10 for each factor, 10 being the common default value. Other safety factors are added to account for use of a LOAEL instead of a NOAEL and to account for inadequacies in the data base.

There are a number of criticisms of the uncertainty factor approach. First, the observed no effect level depends on the sample size. For example, a response of 0 in 10 means something different than a response rate of 0 in 1000. Moreover, the approach assumes that a threshold dose exists below which no adverse effects will occur. These thresholds may vary from animal to animal so that a minimum threshold would need to be estimated.

A benchmark dose approach has been gaining recognition as a method which could address some of these difficulties. This method involves modelling the dose-response curve then estimating the lower 95% confidence limit on the dose corresponding to a small increase in risk over the background rate.

The benchmark dose approach can be applied to either quantal or continuous responses.

Type of data	Model	Model equation
Quantal	**Linear regression**	$P(d)=c+(1-c\{1-\exp[-q_1(d-do)]\}$
	Multistage	$P(d)=1+\exp[-q_1(d-d_o)-q_2(d-d_o)^2- \ldots -q_k(d-d_o)^k]$
Continuous	**Linear regression**	$m(d)=c+q_1(d-d_o)$
	Polynomial regression	$m(d)=c+q_1(d-d_o)+q_2(d-d_o)^2+ \ldots +q_k(d-d_o)^k$

Other curve fitting methods are possible.

For quantal data, the benchmark dose d is that dose with produces a specified extra risk of:

$$\frac{(P(d) \ \& \ P(0))}{(1 \ \& \ P(0))}$$

where $P(0)$ is the value of the model equation $P(d)$ when dose is zero. For continuous data, the benchmark dose is that which produces the specified extra response:

$$\frac{m(d) \ \& \ m(0)}{m(0)}$$

A safety factor is then introduced to bring the predefined benchmark dose from 1% -10%

down to a level of "acceptable" risk. For a more complete discussion of the benchmark dose, see McColl (1990, p.23).

Non-threshold toxicants

For non-threshold toxicants, it is assumed that any exposure carries with it a probability of a response so that the exposure response relationship can be modelled with a smooth curve that approaches background rates as the exposure levels get smaller and smaller. This raises the question of how to express risk.

For evaluation of the Priority Substances under the Canadian Environmental Act (CEPA), the objective of the quantitative risk assessments was to provide a ranking of carcinogens. To do this, the lifetime exposure which yielded a 5% increase in risk was computed directly from the exposure-response relationship. This was combined with estimated exposure of the general Canadian population to obtain an exposure potency index for each substance. The result indices could then be used for ranking.

Other programs such as hazardous wastes and drinking water focus on setting guidelines. For these programs, it is important that enough information is available so that decisions can be made about potential remedial actions. Traditionally, excess risk would be estimated at low doses from the dose response curve.

For both objectives, risk is estimated by modelling the toxic response as a function of exposure. Krewski and VanRyzin (1981) discuss and compare a number of models which fall into two categories. The statistical or tolerance models are based on the assumption that an individual selected at random will respond at date d with probability such that:

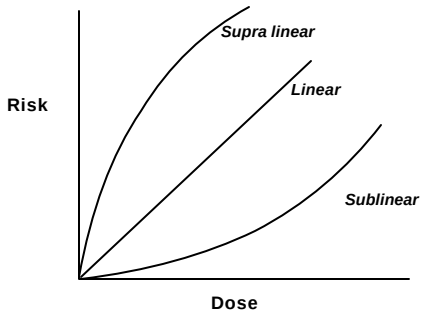
$$P(d) = Pr(\text{ tolerance} \leq d) = F(a + b \log(t))$$

These models include the probit, logistic and extreme value models.

Stochastic or mechanistic models are based on the assumption that a response is induced as a result of the random occurrence of one or more biological events. For instance, for the one-hit model, a response will be induced after the target site bas been hit by a single biologically effective unit of dose within a specified time interval. For the multi-hit model it is assumed that a response may induced after multiple hits. The multi-stage model assumes that the induction of irreversible self-replicating toxic effects is the result of the occurrence of a number of different random biological events, with the age-specific rate of occurrence of each event linearly related to dose.

Other regulatory programs focus on determining cancer risk at ambient levels of exposure. Ambient levels are usually much lower than experimental exposure levels in animal bioassays. Exposure-response models are derived based on these relatively high exposure levels necessitating extrapolation of risks at higher levels to risks at lower levels. However, these models are sensitive to the experimental exposure groups since the shape of the curve at the low levels is unknown.

Dose-Response Shapes at Low Dose



Multistage model

The multistage model is used by the USEPA and other regulatory agencies for estimating

cancer risk. It is based on the probability $P(d)$ of a tumour occurring following an exposure to a dose d at a fixed time t taking the form:

$$P(d) = 1 - \exp[-ct^{k} \sum_{i=0}^{k} (q_i d^{i})]$$

where k is the number of stages assumed in the carcinogenic process. As the dose levels get smaller and smaller, the excess risk becomes approximately linear in dose. The 95% upper confidence limit is then computed and used as the low dose slope factor. The inverse of the slope is called the unit risk. The method is referred to as the Linearized Multistage (LMS) method.

Robust linear modelling

A number of problems were pointed out with extrapolating risks to ambient levels using models developed at relatively higher dose levels. Krewski et al (1991) developed a method (called the *model free extrapolation method* (MFX) or *robust linear modelling)* which addresses some of these concerns. The only assumption required by this method is that the dose-response curve is linear at low doses. Consider a bioassay with t+1 dose groups, including a control group. Estimate the probability of a response at each dose group by $pi = r_i/n_i$ where r_i is the number of animals in the ith dose group with the response and n_i is the number of animals in the ith dose group. Compare the p in the control group to the p in each dose group using Fishers Exact test at the 5% level of significance. Then, calculate the upper 95% confidence limit on p for each dose group before the first significant dose group and the lower 95% confidence limit on the p for the control group using exact binomial intervals and 5/(t+1) as the significance level according to the Bonferoni inequality. Calculate the slope of the line between these upper and lower

confidence limits and select the smallest slope as the risk factor.

The strengths of the MFX method are that no model is assumed that may greatly influence the low dose extrapolation and that a reasonable degree of confidence exists that the true risk is below what MFX produces. Weaknesses include the fact that not all the data is used in the estimation of potency at low doses. Krewski et al (1991) show that the median ratio of the MFX to LMS estimates of the low dose slope was 1.3. Thus MFX gives somewhat higher slopes than LMS. In 443 of the 572 experiments, the ratio was within a factor of two.

The cancer risks estimated using standard methods are based on experimental exposure which may be administered intermittently though consistently throughout the lifetime of the animal. In order to obtain the risk associated with a constant average daily dose in mg/kg bw (body weight in grams)/day, the exposure must be adjusted. For example, say in an inhalation study, mice were exposed to concentrations in air for eight hours a day, 5 days a week for 92 weeks. The standard lifetime of mice is considered to be 104 weeks. To convert the exposure to constant exposure, the intermittent concentration is multiplied by a scaling factor as follows:

constant exposure conc. = experimental conc. x (5/7) x (8/24) x (92/104)

Sometimes the experiment does not end at the standard lifetime of the animal under study. If the experiment ended early, the number of animals with tumours will probably be lower than if the experiment had been allowed to run to standard lifetime duration. In making comparisons with other carcinogens, the potency may seem lower. Experiments suggest that the tumour rate increases more rapidly than linearly with age and thus adjustment of the tumour rate by a factor f^2 or f^3 where:

f = time animals are on test / standard lifetime

is appropriate.

Sometimes exposure to a substance is not constant nor consistent over time. For example, infants and children may be exposed to higher amounts of pesticide residues that may be found in apple juice on a mg/kg bw/day basis than adults. Murdoch, Krewski and Wargo (1992) discuss this issue for the multistage and two-stage birth-death mutation models. They found that approximating lifetime risk on the basis of lifetime average daily dose may underestimate or overestimate risks depending on the exposure pattern.

Animal to human extrapolation

Animals used in toxicology experiments are smaller than humans and usually have higher metabolism rates. Therefore often the body surface area is used as the basis for extrapolation since it is believed to reflect more differences in metabolism than body weight. It is first assumed that the different species are equally sensitive on a basis of dose per unit surface area. A commonly used conversion factor is then:

$$\text{interspecies adjustment factor} = 3(70/w)^{1/2}$$

where 70 kilograms is the standard body weight of a human and w (in kilograms) is the standard animal body weight for the particular species used in the study.

There are other conversion factors which are advocated and can be written using the general form:

$$\text{potency} = a \times (\text{body weight})^{b}$$

where b takes on different values (see table below)

b	Interspecies Extrapolation Based On
1	body weight
2/3	body surface area
3/4	metabolic rate

4.2 Probabilistic assessment

Exposure and risk assessment in HIA is subject to many uncertainties. For example, risk is often estimated based on animal toxicology studies. These studies are conducted according to strict protocols on homogeneous populations of rodents. Extrapolating risk from rodents to humans is highly uncertain. In the absence of mechanistic information to the contrary it is assumed that humans will have the same risk as the rodents. That this is not always the case has been borne out by a number of studies which show that effects displayed in some species of rodents are not observed in others.

Additional sources of uncertainty arise from the extrapolation from the high doses used in laboratory studies to the lower levels in exposure experienced by humans; when toxicological results obtained under one route of exposure (oral, inhalation, dermal) are extrapolated to another; or when results of subchronic studies are used to estimate thresholds for chronic effects. Epidemiological data on exposed human populations can also be subject to considerable uncertainty when retrospective exposure profiles are

difficult to construct, particularly with chronic diseases such as cancer for which exposure data many years prior to disease ascertainment are needed. Even prospective studies have large uncertainties. For example, in dietary studies both recall and food diaries can be subject to error. Disease diagnoses are subject to error as well.

Point estimates of risk in combination with conservative assumptions have historically been the way of expressing risk. However, the point estimates are based on highly uncertain data. By focussing on single summary estimates, it is possible to lose sight of the fact that such estimates are more properly viewed as upper limits rather than best estimates.

A more complete characterization of risk using uncertainty analysis can help us better understand the role of uncertainties in the overall risk and exposure estimates. The results can be summarized in the form of a distribution of possible risks or exposures to take into account as many sources of uncertainty and variability as possible. Morgan, Henrion, and Small (1990), NRC (1994) and Bartlett et al (1996) contain more detailed discussions of uncertainty analysis.

In an example, Hoffman and Hammonds (1994) used a multiplicative risk model to estimate the hazard quotient for a population of individuals potentially exposed through the ingestion of contaminated fish. The risk *R* is expressed as:

$$R = X_1 \times X_2 \times (X_3)^{-1} \times (X_4)^{-1}$$

where X_1 is the concentration of the contaminant in fish, X_2 is the ingestion rate of fish, X_3 is the body mass of a human and X_4 is the reference dose. The overall uncertainty distribution for R is developed by first estimating distributions for the input variables, X_1,

X_2, X_3, X_4, sampling from each of the input variables using Monte Carlo analysis, inputting the sampled values into the risk equation and computing the resulting R. This is repeated thousands of times to develop a sampling distribution for R. Monte Carlo analysis, though convenient and easy to use applying programs like Crystal Ball and @Risk, needs to be applied carefully to ensure the integrity of the results (Burmaster and Anderson, 1994).

Rai et al (1996) stress the importance of distinguishing between inherent variability in the input parameters and uncertainty resulting from a lack of knowledge. In the example above, the authors assume that X_4 is subject to only uncertainty, X_3 to variability and X_1 and X_2 to both uncertainty and variability. Distributions are developed for each component and then applied using Monte Carlo analysis. The impact of the various sources of variability and uncertainty can then be assessed individually or together.

Difficulties, Challenges

There are a number of difficulties and challenges associated with utilizing biostatistics to ensure evidence-based decision making.

- The complexity of the analysis can require an in-depth understanding of statistical concepts in order to be properly utilized or reviewed. The more complex the data acquisition protocol, the more complex statistical procedures needed for analysis.

- Some simple statistical methods can always be applied but they must be interpreted in terms of the way the data was collected.

- The comfort level in chosen statistical methods can be improved by obtaining training in concepts and consulting with experienced statisticians when necessary.

It is not always necessary to consult a statistician; it is more important to understand the concepts enough so as to know when to consult one.

- It is best to consult a statistician who is familiar with the types of statistical methods utilized in HIAs. For example, specific methods are utilized for developing risk factors. An understanding of the limitations in the data upon which these risk factors are based is essential to their appropriate interpretation.

Conclusions

- Statistical methods can enhance the credibility of results.
- They also provide a good indication of variability associated with the data.
- Care must be exercised in selecting methods of analysis and in the subsequent interpretation of the results.
- Knowing when to consult a statistician is helpful. A statistician with experience in HIA/EIA would be most helpful.

Resource materials, persons, organizations

Learning tools: Basic textbooks for non-statisticians, introductory courses, multi-media CD ROM.

There are many fine textbooks available which explain in detail basic statistical concepts. It is best to focus on textbooks written for non-statisticians since those for statisticians tend to be highly technical in

language and presentation. It is highly recommended that a course in statistics be taken before any self education is attempted. Statistics is built on difficult concepts of probability which are easier to learn in an interactive environment.

Courses and textbooks can provide the highlights of statistical concepts but cannot possibly teach the entire field of statistics since the field is so rich. As stated earlier, statistical methods always depend on making assumptions about the data. Statisticians are trained to recognize these assumptions and to chose methods accordingly or develop/modify methods as needed. If the assumptions are ignored, then errors in conclusions could result. Errors range from inconsequential to very serious which could, for example, lead to costly decisions about mitigation where such action may not be justified if the data had been analysed correctly.

Consulting with a statistician in your organization at appropriate times can help avoid these costly mistakes. If a statistician is not available in your organization, there may be some available at the local University or in a consulting firm. The statistician is more valuable if they have developed an understanding of the types of data and issues in HIA. For example, they should have some experience in exposure modelling and toxicological study design.

Many computer programs are also available which perform statistical analyses. Spreadsheet programs such as Lotus 1-2-3, Excel, and Quatropro have regression analyses, plotting capabilities, and other statistical functions in addition to their spreadsheet capabilities. Other programs are designed specifically for statistical analysis (such as SAS, SPlus, SPSS, etc.) Add on programs such as Crystal Ball (Decisioneering, 1996) do simple Monte Carlo analyses. Interactive multimedia CD ROMs provide courses in statistics such as ActivStats (Velleman, 1998) .

References

Bartlett, S., Richardson, G.M., Krewski, D., Rai, S.N., and Fyfe, M. 1996. Characterizing Uncertainty in Risk Assessment - Conclusions Drawn from a Workshop *Human and Ecological Risk Assessment*, **2**, 221-231

Burmaster, D.E. and Anderson, P.D. 1994. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Anal*. **14**, 477-481.

Decisioneering. 1996. *Crystal Ball, Version 4.0 User Manual*. Denver, Colorado, Decisioneering, Inc.

Hoffman, F.O. and Hammonds, J.S. 1994. Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Anal*. **14**, 707-712.

Johnson, R.A., and Wichern, D.W. 1992. Applied Multivariate Statistical Analysis. 3rd ed. Englewood Cliffs, New Jersey, Prentice-Hall, Inc. 642 pages.

Kennedy, John B. and Neville, Adam M. 1974. *Basic Statistical Methods for Engineers & Scientists.* 2nd ed. New York, Harper & Row Publishers, 490 pages.

Krewski, D. and Van Ryzin 1981. Dose response models for quantal toxicity data. in *Statistics and Related Topics*, M. Csorgo, D.A. Dawson, J.N.K. Rao, A.K.Md.E. Saleh (eds.), North-Holland Publishing Company, 1981, 201-231.

Krewski, D., Gaylor, D., and Szyszkowicz. 1991. A model-free approach to low-dose extrapolation. *Environmental Health Perspectives*. **90,** 279-285

MathSoft 1997. *S-Plus 4 Guide to Statistics*. Seattle, Washington, MathSoft, Inc.

McColl, R.S. 1990. Biological Safety Factors in Toxicological Risk Assessment" *Health*

Canada Report Number 90-EHD-154.

Morgan, M., Henrion, M., and Small, M. 1990. *Uncertainty*. New York, Cambridge University Press, 332 pages.

Murdoch, D.J., Krewski, D. and Wargo, J. (1992) "Cancer Risk Assessment with Intermittent Exposure" *Risk Analysis* **12** 569-577

Nete, J., Wasserman, W., Kutner, M.H. 1990. Applied Linear Statistical Models. 3rd ed. Homewood, IL, Richard D. Irwin, Inc. 1181 pages.

Rai, S.N., Krewski, D., and Bartlett, S. 1996. A General Framework for the Analysis of Uncertainty and Variability in Risk Assessment. *Human and Ecological Risk Assessment*, **2**, 972-989

Shoukri, M.M., Edge, V.L. 1996. Statistical Methods for Health Sciences. Boca Raton, CRC Press, 298 pages.

SPSS 1990. *SPSS Advanced Statistics Student Guide.* Chicago, Illinois, SPSS Inc.

Statistical Analysis System (SAS) 1990. *SAS/STAT User's Guide*. Cary, NC, SAS Institute Inc.

Velleman, P. 1998. Reading, MA, Addison Wesley Longman.