

Understanding Epidemiology for Chemical Risk Assessment

Introduction - a brief overview of
epidemiological methods and some
statistical topics



The Interdepartmental Group
on Health Risks from Chemicals

Course aims

- To provide an introduction to epidemiological methods
- To facilitate critical evaluation of published material in which these methods are used
- To give sufficient background to assist in decision and policy making processes

Introduction to epidemiology

- Epidemiology
 - is the study of the distribution of disease in human populations
 - and of factors that determine the distribution
- Objectives
 - to understand the aetiology (causes) of disease
 - to identify preventive measures
 - to evaluate preventive measures and therapeutic interventions

Introduction to epidemiology (cont)

- Descriptive studies
 - summarise distributions of disease incidence, mortality, survival, etc
- Intervention (experimental) studies
 - investigate relationships between disease distribution and other factors in controlled conditions

Introduction to epidemiology (cont)

- Analytical (observational) studies
 - investigate associations between disease distribution and factors that might be involved in aetiology of the disease
 - include
 - cross-sectional (snap-shot) studies
 - cohort (prospective) studies
 - case-control (retrospective) studies

Introduction to epidemiology (cont)

- Issues to consider
 - study objectives
 - target and study populations
 - sample size and sampling method
 - existence of and adherence to a written protocol
 - training of study staff
 - piloting of study methods
 - response rate
 - verification of data

Important statistical concepts

- Target population
 - a group of individuals about whom we seek information
 - eg children in Hammersmith
- Population mean
 - the average value of some characteristic in the population
 - eg the average birth weight of children in Hammersmith
- Population standard deviation
 - a measure of the spread of values (variability) of some characteristic in the population eg the mean

Important statistical concept

- Sample
 - a subset of the target population which is used to estimate population parameters
 - eg 2000 children born in Hammersmith in a specified period
- Sample mean
 - mean calculated from the sample and used as estimate of the population mean
 - eg mean birth weight of 2858 g
- Standard error of the mean
 - a measure of the spread of values (variability) of the sample mean that would be obtained if we were to draw samples of the same size repeatedly
(the sample standard deviation divided by the square root of the sample size)

Important statistical concepts

- Confidence interval for the mean
 - a range of values within which the population mean is expected to lie with a certain confidence
 - the general formula is
$$\text{sample mean} \pm \text{critical value} \times \text{standard error of the mean}$$
 - the critical value depends on the statistical distribution that the sample mean is assumed to follow
 - and the degree of confidence required (typically 95%)
 - eg a 95% confidence interval for the mean birth weight of 2837–2879 g

Important statistical concepts

- Interpretation
 - 95% of samples selected in the same way as our particular sample would produce a 95% confidence interval that would contain the population mean
 - ie we are using a method that works 95% of the time (and fails 5% of the time)

Important statistical concepts

- Hypothesis testing
 - a framework for drawing conclusions about a target population from a sample
- Null hypothesis (H_0)
 - the hypothesis to be tested
 - typically the negation of the hypothesis that really interests us
 - eg the mean birth weight of children in Hammersmith is 2850 g
- Alternative hypothesis (H_1)
 - summarises departures from the null that interest us
 - eg the mean birth weight of children in Hammersmith is not 2850 g

Important statistical concepts

- Test statistic
 - a measure of the strength of departure from the null hypothesis
 - eg a statistic based on the difference between the sample mean and the mean under the null hypothesis
- Probability (p value)
 - calculate the probability of obtaining a test statistic as extreme or more extreme than the one for our sample if the null hypothesis is true
 - eg $p=0.59$ in the test for comparing the sample mean birth weight of 2858 g with the hypothesised mean of 2850 g

Important statistical concepts

- Use the probability to decide whether or not to reject the null hypothesis
- Large probability $\equiv$ event occurred which is not uncommon $\Rightarrow$ no reasons to reject H_0
- Small probability
 - either H_0 true and improbable event occurred
 - or H_0 is false $\Rightarrow$ reject H_0
- Significance level is the value of the probability at which it is decided that the event is sufficiently improbable to reject H_0

Important statistical concepts

- Interpretation
 - traditionally, significance levels less than 0.05 are regarded as statistically significant and in such cases the null hypothesis is rejected in favour of the alternative hypothesis (at the 5% level)
 - but with $p=0.59$ we cannot reject the null hypothesis that the mean birth weight of children in Hammersmith is 2850 g

Important statistical concepts

- Confidence intervals revisited
 - a 95% confidence interval that excludes the mean value of the test statistic under the null hypothesis represents a difference that is significant at the 5% level
 - but with a 95% confidence interval of 2837–2879 g, which includes the value 2850 g, we cannot reject the null hypothesis that the mean birth weight of children in Hammersmith is 2850 g

Important statistical concepts

- Power of a study
 - The probability of detecting a real difference between groups i.e. the probability of rejecting the null hypothesis when the alternative hypothesis is true
 - depends on various factors, including the sample size (bigger studies generally have greater power)