Statistical Notes: An introduction

RK McKinley, A Sheikh and A Cook

Statistics is an obscure and threatening discipline for many clinicians. Nevertheless, an understanding of the principles of statistics is required when evaluating research papers and appraising the evidence on which our practice should be based. Furthermore, although statistical advice is increasingly available to novice researchers, statisticians are fallible. No one understands a project better than the researcher whose responsibility is to decide if the advice given is appropriate. This requires some understanding of statistics. Therefore, a lack of familiarity with statistics is a barrier to those who would like to undertake research and to use its results.

The basics of statistics are really quite straightforward. At the simplest level, statistics is about summarising and presenting data in ways which accurately reflect their meaning. The next level is 'hypothesis testing', when the researcher wants to know whether the 'treatment worked' or whether a possible risk factor, say birth weight, is associated with a particular outcome, such as the likelihood of developing asthma. The essence of this is to determine the likelihood of the differences between groups arising by chance. If, in a well designed study, it is unlikely that the study groups are similar the researcher has 'got a result' (although a 'negative' result can be just as important). This illustrates an aspect of statistics which is often forgotten – its contribution to study design. Finally, closely related to hypothesis testing is estimation, whereby the researcher goes on to investigate just how different the groups are.

The aim of this series is to introduce the novice to basic statistical concepts using examples from the respiratory research literature. We do not claim to provide a comprehensive introduction, rather to provide the reader with sufficient understanding and vocabulary to appraise the statistics used in a research paper or start to discuss his or her research ideas or projects with a statistician.

The seven articles (see Box) in the series will take the reader through data presentation (numbers I and II), the principles of hypothesis testing (number III), estimates and confidence limits (number IV), both epidemiological and intervention study design including the principles of obtaining a study sample (number V), sample size calculations (number VI) and a summary which will examine the some of the 'nuts and bolts' of statistics.

We believe that an understanding of these basics will provide a valuable introduction to statistics. These notes will enable clinicians to better understand published research and lower the barriers to embarking on research for those who wish to start. ∎

SERIES OUTLINE

I Descriptive statistics (Part 1)
Simple but important errors are common in the choice of means, medians, SDs and ranges to summarise data which can be avoided by the application of simple rules.

II Descriptive statistics (Part 2)
This article will discuss more sophisticated ways of making information accessible to the reader such as relative risks, absolute differences, odds ratios and numbers needed to treat.

III Hypothesis testing
The limitations and interpretation of tests of hypothesis will be discussed using t-tests and chi-square tests as exemplars.

IV Estimation
Standard errors and confidence intervals and their dependence on sample size and variability will be described together with the relationship between p-values and confidence intervals.

V Study design
A discussion of retrospective and prospective study designs and the importance of control groups and randomisation in study design.

VI Sample size and power
The importance and principles of calculating sample sizes will be discussed together with the weaknesses of small studies. This Note will be of practical relevance to readers who are planning to conduct research.

VII Practicalities
Practical problems relating to the use of computers and commercially available statistical packages together with their strengths and weaknesses will be discussed. Coding of data, handling missing values, and transferring data between packages will also be addressed.

Robert McKinley
Senior Lecturer

Department of General Practice and Primary Health Care, University of Leicester, Leicester General Hospital, Gwendolen Road, Leicester LE5 4PW, UK

Aziz Sheikh
Clinical Research Fellow

Adrian Cook
Statistician

Department of General Practice and Primary Health Care, Imperial College School of Medicine, Norfolk Place, London W2 1PG, UK

Correspondence to:
Dr A Sheikh
aziz.sheikh@ic.ac.uk