PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2005; **4**: 129–139
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pst.168

Web-based resources to assist the statistical analysis and presentation of data

MAIN PAPER

Mark A. Mullee*,†
Research & Development Support Unit, University of Southampton, Southampton, U.K.

The intention of this article is to highlight sources of web-based reference material, courses and software that will aid statisticians and researchers. The article includes websites that: assist in writing a protocol or proposal; link to online statistical textbooks; and provide statistical calculators or links to free statistical software and other guidance documents. Copyright © 2005 John Wiley & Sons, Ltd.

Keywords: ***website; statistical textbook; statistical calculator; statistical software***

1. INTRODUCTION

When working day to day as an applied statistician it will often be the case that questions are asked by colleagues or you wish to obtain a quick answer to a problem. The World Wide Web, books and journals offer a vast array of resources to support both the statistician and other researchers in the analysis and presentation of data. This article will guide the reader through some of the web-based resources, grouped according to subject.

Although this article cannot profess to be definitive and cannot vouch for the accuracy or validity of the material highlighted, the websites recommended in this article are ones used by the author or colleagues in their consulting and teaching. All the websites (and links) quoted in the paper were last accessed on 21 March 2005.

*Correspondence to: Mark A. Mullee, RDSU, Level C (805), South Academic Block, Southampton General Hospital, Southampton SO16 6YD, U.K.
†E-mail: m.a.mullee@soton.ac.uk

2. GUIDELINES FOR WRITING DISSERTATIONS IN MICROSOFT WORD

The University of Stirling Information Services have produced some invaluable advice for those who use Microsoft Word to write a dissertation or other lengthy documents (http://www.is.stir.ac.uk/documentation/Guide_to_Word-Processing_Long_Documents.pdf). The advice covers, the application of styles and how to generate tables of contents, as well as lists of figure and tables. It has an easy-to-follow style which demonstrates each application through example. I recommend this document to all students and also to anyone

using Word who wishes to use any of its functionality.

3. STATISTICAL CONCEPTS

There are many sources of reference materials that offer assistance in explaining statistical concepts. The following is a selection of useful resources. Sample size and power concepts and calculators will be dealt with in a separate section.

3.1. StatSoft

StatSoft's Electronic Statistics Textbook (http://www.statsoft.com/textbook/stathome.html) offers definitions and training in the understanding and application of statistics. It starts with an explanation of elementary statistical concepts, and then offers detailed explanations of specific topics (Figure 1). A useful feature of the in-depth explanations is the use of diagrams and animated figures. The statistical advisor uses answers to successive questions about the nature of the research to suggest which statistical methods should be used and where to find them. This resource also includes a comprehensive statistical glossary.

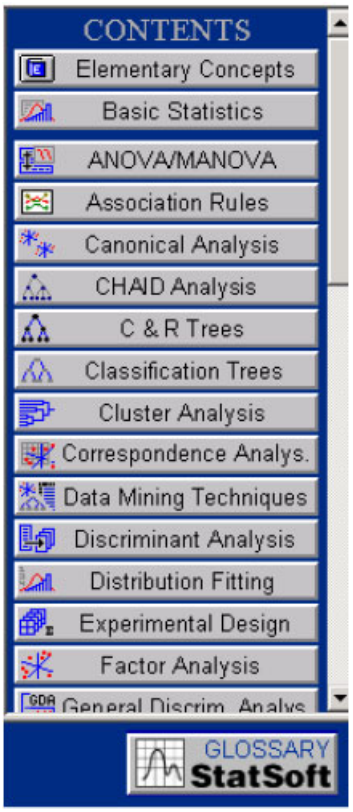

Figure 1. The Electronic Statistics Textbook.

3.2. Statistics Notes in the *British Medical Journal*

Two statisticians (Doug Altman and Martin Bland) have edited a list of a series of statistics notes published by the *British Medical Journal*. The series is listed, with links to the original articles, on Martin Bland's homepage at http://www-users.york.ac.uk/%7Emb55/pubs/pbstnote.htm. The series includes articles describing:

- Treatment allocation in controlled trials: why randomise?
- How to randomise
- Concealing treatment allocation in randomised trials
- Blinding in clinical trials and other studies
- Analysing controlled trials with baseline and follow up measurements
- Transformations, means, and confidence intervals
- Detecting skewness from summary information
- Logarithms
- Bayesians and frequentists.

3.3. Steps

The STEPS consortium has developed problem-based modules to support the teaching of statistics in biology, business, geography and psychology. The software is freely available to educational institutions, and can be downloaded from http://www.stats.gla.ac.uk/steps/index.html. The STEPS software comes with a built-in glossary of statistical terms, in order to provide a supportive educational environment. A web version of the

glossary is also available at http://www.stats.gla.ac.uk/steps/glossary/index.html. This includes definitions of, for example, hypothesis testing (one-sided versus two-sided), confidence intervals, and nonparametric methods.

3.4. Statistics at Square One

An electronic version of the elementary statistics textbook *Statistics at Square One* [1] is available on the BMJ website (http://bmj.bmjjournals.com/collections/statsbk/index.shtml). This textbook introduces topics from data display and summary to survival analysis.

3.5. Betty C. Jung's Website

Although it has an American public health focus, Betty C. Jung's Website (http://www.bettycjung.net) groups together a vast array of useful links under a number of headings. The aim of the site is 'to provide a place where public health and healthcare professionals, and anyone interested, can find good quality information that exist on the World Wide Web'.

Selecting Biostatistics from the Popular Pages reveals a choice of topic groups with links to further websites:

Biostatistics/Statistics Sites
Charting & Graphing Data
Public Health Statistical & Mapping Software
Software Downloads
Statistical Procedures
Statistical Programs

Choosing Charting & Graphing Data, for example, presents the user with a huge choice of links grouped into 15 categories, including the following:

General Resources
Charting Data
Cumulative Frequencies
Frequency Distributions
Graphing
Graphing with Excel
PowerPoint Help
Quality Control Graphing

Clicking on Excel Charting Tutorials under Charting Data takes one to a valuable site for users of Microsoft Excel.

Links accessed by clicking on Statistical Procedures are likewise grouped under various headings, for example:

ANOVA
Contingency Tables
Data Collection
Online Textbooks/Tutorials
Rater Agreement
Regression Analysis
Sample Size Information

Useful definitions of statistical significance and confidence intervals can be found here. There is also a link to StatPrimer (version 6.1), which has electronic versions of probability tables (z, t, χ^2, etc.).

Statistical Programs offers links related to statistical, spreadsheet and database software, including SAS, Stata, SPSS, Epi Info, Microsoft Excel and Microsoft Access.

4. STATISTICAL SOFTWARE

4.1. StatPages.net

The web pages listed at http://members.aol.com/johnp71/javastat.html are described as comprising 'a powerful, conveniently-accessible, multi-platform statistical software package. There are also links to online statistics books, tutorials, downloadable software, and related resources. All of these resources are freely accessible, once you can get onto the Internet'. Figure 2 is a snapshot of part of the home page.

Whilst there are links to free statistical software packages and add-ins (e.g. for Microsoft Excel), the strength of this site lies in its links to statistical calculators. To take just one example, the Contingency tables, cross-tabs, Chi-Square tests link reveals an extensive list of calculators, including


Sections of the *StatPages.net* web site

Interactive Stats | Free Software | Books & Manuals | Demo's & Tutorials | Other Links | About this Website | What's New | My Home Page

Web Pages that Perform Statistical Calculations!

(StatPages.net)

Over 600 Links (including 330 Calculating Pages) -- And Growing!

(Updated 04/28/2005 -- check out What's New, and the Awards and Recognition this site has received.)

The web pages listed here comprise a powerful, conveniently-accessible, multi-platform statistical software package. There are also links to online statistics books, tutorials, downloadable software, and related resources. All of these resources are freely accessible, once you can get onto the Internet. FREE dial-up Internet access is available from NetZero and Juno.

These pages are located on servers all over the world, and are the result of much cleverness and hard work on the part of some very talented individuals. So if you find a page useful it would be nice to send the authors a short e-mail expressing your appreciation for their hard work and generosity in making this software freely accessible to the world. Please let me know of any dead links, computational errors, or other problems you might encounter (e-mail me at johnp71@aol.com).

Table of Contents for this page...

- Selecting the right kind of analysis
- "Online Software" Package websites
- Calculators, plotters, function integrators, and interactive programming environments
- Probability distribution functions: tables, graphs, random number generators
- Descriptive statistics, histograms, charts
- Confidence intervals, single-population tests
- Sample comparisons: t-tests, ANOVAs, non-parametric comparisons
- Contingency tables, cross-tabs, Chi-Square tests
- Regression, correlation, least squares curve-fitting, non-parametric correlation
- Analysis of survival data
- Bayesian Methods
- Other statistical tests and analyses
- Specialized and discipline-specific tests and analyses
- Power, sample size and experimental design

Figure 2. Part of StatPages.net front page.

one which calculates chi-square, sensitivity, odds ratio, relative risk, difference in proportions, number needed to treat, and many others, all with confidence intervals. I find this useful if I wish to perform a quick calculation but do not wish to go through the full process of launching SAS or SPSS.

4.2. Epi Info™

Epi Info™ may be downloaded free from http://www.cdc.gov/epiinfo/. Epi Info™ is software written by the Centers for Disease Control and Prevention for epidemiologists and other public health and medical professionals to develop a questionnaire or form, customize the data entry process, and enter and analyse data. Epidemiological statistics, tables, graphs, and maps can be produced. Epi Map displays geographic maps with data from Epi Info™. Researchers in the pharmaceutical industry may not feel the necessity to use this package, but a lot of researchers do use it and write journal articles on studies that have been data-managed and analysed using this system.

4.3. UCLA website

The Statistical Computing Resources website (http://www.ats.ucla.edu/stat/) has some excellent material and is a site to which I personally often refer. Its front page is shown in Figure 3.

For those wishing to learn more, need a refresher or have questions about using statistical packages, the Statistical Computing Packages section offers help, for example on SAS, Stata, SPSS, MLwiN and S-Plus. The starter kits for SAS, Stata and SPSS are particularly impressive, with options to view animated instructions with sound (Class notes with movies) on how to use these packages. These animations are very useful, no matter how experienced the user may be. It is great to be able to sit back and listen to a presentation on a given topic at your desk.

Figure 3. UCLA Statistical Computing Resources front page.

What Statistical Analysis Should I Use? (Statistics section) not only suggests which statistical test is appropriate but also instructs how to perform the test using SAS, Stata or SPSS, and how to interpret the results.

The Statistics Papers for Download area provides access to specific papers (e.g. [2–4]) and links to sites with further lists of downloadable papers. The papers are grouped into subject areas, including:

Assumptions
Logistic Regression
Missing Data
Regression
Survey Research
Survival Analysis

4.4. Randomisation Software and Services

There are organizations and software (some free) for performing randomization in clinical trials. The Directory of Randomization Software and Services website (http://www.sghms.ac.uk/depts/phs/guide/randser.htm) has a comprehensive list of these resources. These include the three free online random generators at *Randomization.com*; and Randomizer (a web-based randomization package from the University of Graz at http://www.randomizer.at). There are also links to randomization services which provide, for example, trial support services including telephone randomization. You may not be able to use these packages routinely for industrial trials, but they are useful nonetheless.

5. FREE SAMPLE SIZE AND POWER SOFTWARE

The following is a selection of software available for sample size and power calculations. If you were to use any of these packages, then it is recommended that you use them in parallel with published sources [5,6].

Pharmaceutical STATISTICS

5.1. PS

PS: Power and Sample Size Calculation can be downloaded from (http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize).

> The program runs on the Microsoft Windows operating systems (Windows 95 and later). It can be used for studies with dichotomous, continuous, or survival response measures. The alternative hypothesis of interest may be specified either in terms of differing response rates, means, or survival times, or in terms of relative risks or odds ratios. Studies with dichotomous or continuous outcomes may involve either a matched or independent study design. The program can determine the sample size needed to detect a specified alternative hypothesis with the required power, the power with which a specific alternative hypothesis can be detected with a given sample size, or the specific alternative hypotheses that can be detected with a given power and sample size.
>
> The PS program can produce graphs to explore the relationships between power, sample size and detectable alternative hypotheses. It is often helpful to hold one of these variables constant and plot the other two against each other. The program can generate graphs of sample size versus power for a specific alternative hypothesis, sample size versus detectable alternative hypotheses for a specified power, or power versus detectable alternative hypotheses for a specified sample size. Linear or logarithmic axes may be used for either axis. Multiple curves can be plotted on a single graphic.

The study designs that can be evaluated are: continuous response measures in two groups (independent or paired t tests); survival studies (log-rank tests); linear regression (one or two treatments); case–control studies; matched case–control studies; and cohort studies with dichotomous outcomes.

5.2. Russ Lenth's Java applets for power and sample size

As well as running from the internet (http://www.stat.uiowa.edu/~rlenth/Power/), Russ Lenth's applets may be downloaded to run on a PC. Sample size or power calculations may be calculated for: CI for one proportion; test for one proportion; test comparing two proportions; CI for one mean; one-sample t test (or paired t test); two-sample t test; balanced ANOVA; chi-square test; Poisson test. Each calculation provides a graphical interface for studying the power of one or more tests. They include sliders (convertible to number-entry fields) for varying parameters, and a simple provision for graphing one variable against another. Figure 4 shows a dialogue completed for a two-sample t test.

6. EVIDENCE-BASED MEDICINE

6.1. Centre for Evidence-Based Medicine

The Centre for Evidence-Based Medicine (EBM) website (http://www.cebm.net/) provides links to pages on:

- Learning EBM – defining EBM, background issues, glossary, study designs. These include PowerPoint presentations (e.g. An Introduction to EBM by Martin Dawes) as well as text and links.
- Doing EBM – asking well-built clinical questions, journal searching, critical appraisal (process of deciding whether a piece of research can help in answering the clinical question).
- Teaching EBM – courses, downloads (critical appraisal tools, statistical tools).
- EBM Toolbox – number needed to treat, likelihood ratio (definition and application), pre-test probabilities (definition and application).

6.2. Bandolier

This internet journal (http://www.jr2.ox.ac.uk/bandolier/) is about healthcare, using evidence-based medicine techniques to provide advice about particular treatments or diseases for healthcare professionals and consumers.

6.3. Cochrane Collaboration

This international organization (http://www.cochrane.org/) aims to help people make well-informed decisions about healthcare by preparing, maintaining and promoting the accessibility of

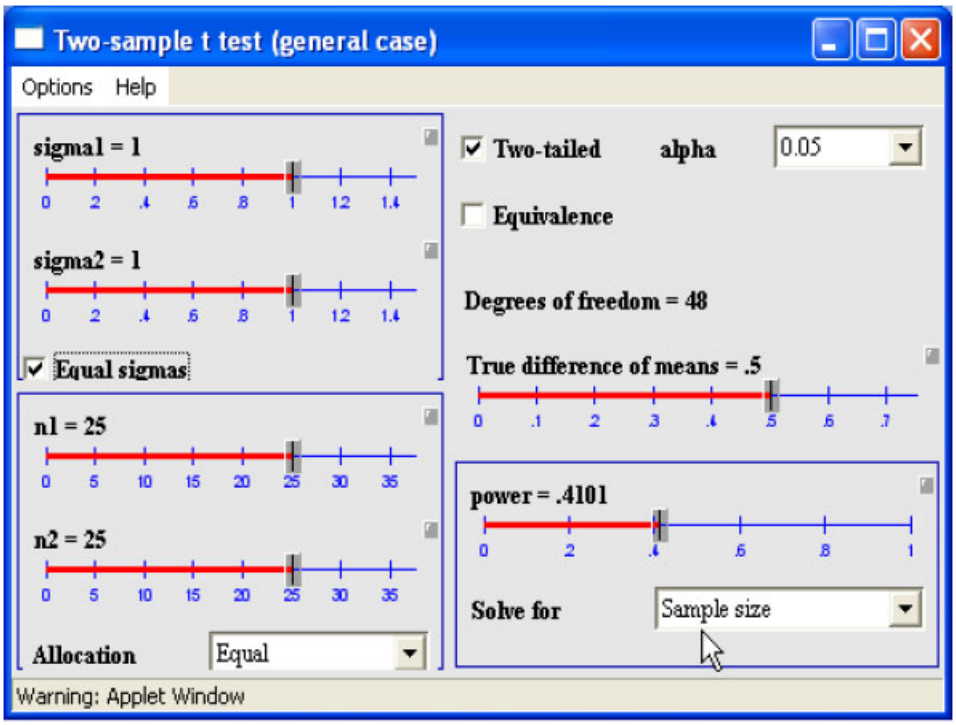

Figure 4. Russ Lenth's applet for a two-sample *t* test.

systematic reviews of the effects of healthcare interventions. The Cochrane Library (http://www.cochrane.org/reviews/clibintro.htm) consists of a regularly updated collection of evidence-based medicine databases, including the Cochrane Database of Systematic Reviews. It is possible to browse and search abstracts of reviews free of charge.

7. REGULATORY GUIDANCE DOCUMENTS

An article could be written entirely about regulatory guidance documents alone, such is their comprehensiveness. I will therefore keep this review brief and highlight the main resources.

7.1. Food and Drug Administration

The US Food and Drug Administration (FDA) provides guidance on aspects of clinical trials. Its website (http://www.fda.gov/cder/guidance/index.htm) lists the guidance documents, including its Good Guidance Practices, and a comprehensive list is also available for download. Links to pharmacology/toxicology guidance documents are also available on this site.

7.2. European Medicines Agency

The European Medicines Agency (EMEA) coordinates the evaluation and supervision of medicinal products throughout the European Union. Regulatory guidance and additional information are available from http://www.emea.eu.int/index/indexh1.htm. Statistical guidelines here include those on missing data, baselines and covariates, multiplicity and non-inferiority margins. The guidelines for different clinical areas and conditions also have statistical guidance included, referring to the most appropriate endpoint(s) and design.

7.3. International Conference on Harmonisation

The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), brings together the regulatory authorities of Europe, Japan and the United States and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration. The official ICH website (http://www.ich.org) includes links to ICH Guidelines, such as *Statistical Principles for Clinical Trials* (obtained through clicking on Guidelines and then on Efficacy) which I assume is very familiar to the readers of this journal. ICH guidelines can also be obtained from the EMEA website.

8. CONSORT

The CONSORT statement (http://consort-statement.org/) takes an evidence-based approach to improve the quality of reports of randomized trials. CONSORT comprises a checklist and flow diagram to help improve the quality of reports of randomized controlled trials. It offers a standard way for researchers to report trials. The checklist includes items, based on evidence, that need to be addressed in the report; the flow diagram is designed to provide readers with a clear picture of the progress of all participants in the trial, from the time they are randomized until the end of their involvement. The intent is to make the experimental process clearer, whether flawed or not, so that users of the data can more appropriately evaluate its validity for their purposes. The CONSORT statement, as published in a number of journals, can be downloaded (http://consort-statement.org/annals.pdf). An explanation and elaboration of the statement is also available (http://consort-statement.org/ene.pdf). The extension of CONSORT to cluster randomized trials is available at http://consort-statement.org/Cluster/cluster.htm.

Possibly, without realizing it, many readers will be familiar with the CONSORT format for reporting as it has been adopted by many of the leading clinical journals.

9. MICROSOFT EXCEL

There are a host of websites that offer advice, tutorials and support in using Excel. The University of Reading provides comprehensive documentation to assist beginners, intermediate and advanced users (http://www.rdg.ac.uk/ITS/info/help/quickhelp/spread.htm); the University of Leeds also provides introductory documentation (http://www.leeds.ac.uk/iss/documentation/ss.html).

Microsoft's educational material to support the use of Excel can be found at http://www.microsoft.com/education/Excel.aspx. There are links to Excel tutorials at http://math.about.com/od/excel/.

The University of Reading Statistical Services Centre's Statistical Good Practice Guidelines include tips and warnings on the use of Excel for statistics (http://www.ssc.rdg.ac.uk/publications/guides/topxfs.html).

XLStatistics (http://www.deakin.edu.au/~rodneyc/XLSTATS.HTM) is a set of Microsoft Excel workbooks (for version 97 and later) intended to replace and enhance the tools provided with Excel's Data Analysis Toolbox add-in. The workbooks are designed to implement a step-by-step guide to the statistical analysis of data. The University Reading Statistical Services Centre's statistical add-ins for Excel can be found at http://www.ssc.rdg.ac.uk/software/software.html, and there are further links to add-ins at http://members.aol.com/johnp71/javasta2.html#Excel.

10. PREPARING A PROTOCOL/PROPOSAL

Good research starts with a good question. Having decided on a general area of interest, the investigator must determine specific aims and hypotheses, that is, be clear about what is the purpose of the research.

The Statistics Guide for Research Grant Applications is an online handbook written by the statisticians JM Bland, BK Butland, JL Peacock, J Poloniecki, F Reid, P Sedgwick (http://www.sghms.ac.uk/depts/phs/guide/guide.htm). Although the guide is aimed at writing a grant proposal, the points raised can be generalized to protocols for individual clinical trials and, as such, it provides a good reference document for anyone writing a protocol or proposal.

The aim of the handbook is to help applicants to appreciate some of the statistical pitfalls that await them when constructing a grant proposal (or protocol). The authors stress that

> The handbook is not designed to teach statistics but to provide extra information to those who already have a basic statistical knowledge. ... it is hoped that the handbook will make grant applicants more aware of the right questions to ask and the right information to take along to a statistical consultation and, in addition, help them to understand any advice given.

Another aim of the handbook is to try and clarify the sort of checklist that a statistician might use in the process of reviewing a grant proposal. The handbook covers the following topics:

- Describing the study design – e.g. types of study including cohort, case–control, cross-sectional, randomized controlled clinical trials, confounding, interaction, types of data, validity and reliability.
- Clinical trials – eligibility criteria, randomization, parallel groups, crossover, data monitoring committees, when a trial should be stopped early, informed consent, protocol violation and non-compliance.
- Observational studies – choice of control in case–control studies, assessment bias, recall bias, selecting a representative sample, generalizability and extrapolation of results, maximising response rates.
- Sample size calculation – information required to calculate a sample size, explanation of statistical terms (null hypothesis, *P*-value, power, etc.), allowing for losses to the sample, consistency with study aims and statistical analysis.
- Describing the statistical methods – level of detail required, appropriateness of the proposed method, paired and unpaired comparisons, transformations, multi-level data, multiple testing, regression to the mean, intention to treat, cluster randomized trials, collapsing variables, confidence intervals.
- General – statistical software, ethics, critical appraisal, research governance, data protection.

RDDirect (Department of Health http://www.rddirect.org.uk) is a signposting service for researchers. The resources are specifically designed to support health and social care settings researchers within the NHS, but has advice which is of more general appeal. This website provides a top-level summary of the basics of forming and investigating a research question and provides a good overview for consideration. The site gives a useful research process flowchart (http://www.rdinfo.org.uk/flowchart/Flowchart.html) that maps the different stages of research and provides hints, tips and checklists for researchers. It is possible to download a pdf file which includes not only the flowchart but also tips and advice on the following topics:

Turn your idea into a research question
Review the literature
Design the study and develop methods
Writing your research proposal
Issues about funding
Obtain ethical and Trust approval
Collect and collate data
Analyse the data (quantitative and qualitative) and interpret findings
Implications of your research for clinical practice and identifying how findings could be put into practice
Report the study and disseminate findings

Further guidance for research in the NHS is given in the document 'Guidance for Researchers on the New NHS R&D Application Form', produced by the NHS Research & Development Forum. The link is to be found in the Standard

Pharmaceutical STATISTICS

NHS R&D Application Form Documents section of the website (http://www.rdforum.nhs.uk/library.htm).

11. OTHER WEBSITES/RESOURCES

11.1. Data sets available from the National Institutes of Health (NIH) in the USA

Access to data sets from clinical trials is available from:

- National Institute of Mental Health (http://www.nimh.nih.gov/studies/datasets.cfm)
- National Institute of Allergy and Infectious Disease (http://www.niaid.nih.gov/research/aidsdata.htm)
- National Heart, Lung and Blood Institute (http://www.nhlbi.nih.gov/resources/deca/default.htm)

These data are available under the US Freedom of Information Act. To obtain these trial data sets you need to apply to the NIH (details available on the above websites). As part of this application you are required to obtain approval for your proposal from the ethics committee that governs your research. Clinical colleagues may able to guide statisticians who have not been involved in an ethics submission.

11.2. Drug submissions to the FDA

In the USA, the FDA obtains all the data for a drug submission and reanalyses them, and different groups (statisticians, clinicians, etc.) write reports. These reports are available on the web, under the Freedom of Information Act, at http://www.accessdata.fda.gov/scripts/cder/drugsatfda/. The drugs are ordered by US trade names. You need to: click on the drug name; select NDA link (under Drug Name and FDA Application Number); select Approval History and Related Documents; and, in the approval history table, click on a review. The clinical and statistical reviews are particularly useful for investigating possible study designs and endpoints. These reports can also assist in study design by providing information on standard deviations, completion rates, etc. Not only does the FDA website detail drug submissions, it also has lots of very useful information, including minutes of meetings and statement and recommendations.

11.3. WHO Statistical Information System

The WHO Statistical Information System (http://www3.who.int/whosis/menu.cfm) is the guide to health and health-related epidemiological and statistical information available from the World Health Organization (e.g. Statistics by country or region). Most WHO technical programmes make statistical information available, and they will be linked from here. You can also search by keywords within the WHOSIS or throughout the entire WHO site.

11.4. ClinicalTrials.gov

ClinicalTrials.gov (http://www.clinicaltrials.gov/) provides regularly updated information about US federally and privately supported clinical research using human volunteers. It provides information about a trial's purpose, who may participate, locations, and phone numbers for more details. In addition, this site provides resource information explaining and describing clinical trials (click on Understanding Clinical Trials).

11.5. Articles published in the *British Medical Journal*

Details of all relevant articles published in the *British Medical Journal* since January 1998 can be obtained from http://bmj.bmjjournals.com/collections/ by clicking on the desired specialty or topic.

11.6. Office for National Statistics

National Statistics Online (http://www.statistics.gov.uk/) is the home of official UK statistics, reflecting Britain's economy, population and society at national and local level. Summaries and detailed releases are published free of charge. There are links to various mortality and demographic data broken down by region.

11.7. Google search engine for academic pages and journal articles

Google has a search engine dedicated to academic pages and journal articles (http://scholar.google.com/). It is a very efficient way of doing a literature review. If you find an article of interest, the search takes you directly to the journal article for downloading (assuming you have an appropriate subscription to the journal). It also allows you to list (and link directly to) the articles that have cited a particular publication. At time of writing, this search engine was only a beta test version, but I assume it will be upgraded soon.

12. ADDITIONS

The selections of resources referenced in this article represent work in progress. The websites can be accessed from http://www.rdsu.soton.ac.uk/downloads/. I welcome comments and details of additional web-based resources.

ACKNOWLEDGEMENTS

Steven Julious, Paul Strike and Trevor Bryant provided valuable advice and/or details of some of the websites described in this article.

REFERENCES

1. Swinscow TDV, Campbell MJ. *Statistics at square one*. BMJ Publishing Group: London, 1996.
2. Diehr ZFP, Peterson A, McLerren D. The importance of the normality assumption in large public health data sets. *Annual Reviews of Public Health* 2001; **22**:167–187.
3. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statistician. *Statistics in Medicine* 2000; **19**: 1141–1164.
4. Greenland S, Morgenstern H. Confounding in health research. *Annual Reviews of Public Health* 2001; **22**:189–212.
5. Julious SA. Tutorial in Biostatistics: Sample sizes for clinical trials with normal data. *Statistics in Medicine* 2004; **23**:1921–1986.
6. Machin D, Campbell MJ, Fayers P, Pinol A. *Statistical tables for the design of clinical studies* (2nd edition). Blackwell Scientific Publications: Oxford, 1997.

