Research

AUSTRALASIAN PSYCHIATRY

Getting started in research: systematic reviews and meta-analyses

Australasian Psychiatry
2015, Vol 23(1) 16–21

DOI: 10.1177/1039856214562077
apy.sagepub.com
SAGE

Stephen Kisely Committee for Research, Royal Australian and New Zealand College of Psychiatrists (RANZCP), Melbourne, VIC, and; University of Queensland School of Medicine, Wooloongabba, QLD, Australia

Alice Chang Committee for Research, RANZCP, Melbourne, VIC, Australia

Jim Crowe Committee for Research, RANZCP, Melbourne, VIC, Australia

Cherrie Galletly Committee for Research, RANZCP, Melbourne, VIC, Australia

Peter Jenkins Committee for Research, RANZCP, Melbourne, VIC, Australia

Samantha Loi Committee for Research, RANZCP, Melbourne, VIC, and; Academic Unit for Psychiatry of Old Age, Saint Vincent's Health, Saint George's Hospital, University of Melbourne, Melbourne, VIC, Australia

Jeffrey C Looi Committee for Research, RANZCP, Melbourne, VIC, and; Academic Unit of Psychiatry and Addiction Medicine, Australian National University Medical School, Canberra, ACT, and; ACT Health Directorate, Mental Health Service, Canberra Hospital, Woden, ACT, Australia

Matthew D Macfarlane Committee for Research, RANZCP, Melbourne, VIC, and; Graduate School of Medicine, University of Wollongong, Wollongong, NSW and; Illawarra Shoalhaven Local Health District, NSW, Australia

Ness McVie Committee for Research, RANZCP, Melbourne, VIC, Australia

Stephen Parker Committee for Research, RANZCP, Melbourne, VIC, Australia

Brian Power Committee for Research, RANZCP, Melbourne, VIC, and; School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, WA, and; South Metropolitan Area Health Service, Perth, WA, Australia

Dan Siskind Committee for Research, RANZCP, Melbourne, VIC, and; School of Medicine, University of Queensland, Wooloongabba, QLD, Australia

Geoff Smith Committee for Research, RANZCP, Melbourne, VIC, and; WA Centre for Mental Health Policy Research, School of Psychiatry and Clinical Neurosciences, University of Western Australia, Perth, WA, and; Centre for Research into Disability and Society, Curtin University, Perth, WA Australia

Sally Merry Committee for Research, RANZCP, Melbourne, VIC, Australia; and Psychological Medicine, University of Auckland, Auckland, New Zealand

Stephen Macfarlane Committee for Research, RANZCP, Melbourne, VIC, and; Aged Psychiatry Service, Alfred Health, Caulfield, VIC, Australia

Abstract

Objectives: Systematic reviews are one of the major building blocks of evidence-based medicine. This overview is an introduction to conducting systematic reviews and meta-analyses.

Conclusions: Systematic reviews and meta-analyses of randomised controlled trials (RCTs) represent the most robust form of design in the hierarchy of research evidence. In addition, primary data do not have to be collected by the researcher him/herself, and there is no need for approval from an ethics committee. Systematic reviews and meta-analyses are not as daunting as they may appear to be, provided the scope is sufficiently narrow and an appropriate supervisor available.

Keywords: evidence-based medicine, meta-analysis, research, systematic review, statistics, training

Systematic reviews are crucial to evidence-based medicine. Until recently, most were narrative (qualitative) integrations of research findings. While these can be useful, there are concerns about potential subjectivity in selecting and interpreting studies; therefore,

Corresponding author:
Stephen Kisely, School of Medicine, University of Queensland, MacGregor Building, Number 64, Wooloongabba, QLD 4072, Australia.
Email: s.kisely@uq.edu.au

narrative reviews have largely been replaced by systematic reviews that, if possible, integrate research findings statistically. A systematic review is a critical assessment and evaluation of all research studies on a particular issue. It entails an organised method of locating, assembling, and evaluating the relevant literature, using a set of specific criteria.

A systematic review typically includes a description of the collected findings and may also include a quantitative pooling of data, called a meta-analysis.[1] Systematic reviews and meta-analyses of randomised controlled trials (RCTs) represent the most robust form of design in the hierarchy of research evidence (Figure 1). In addition, primary data do not have to be collected by the researcher him or herself, and so there is no need for approval from an ethics committee; however, supervision is advisable when undertaking a review for the first time.

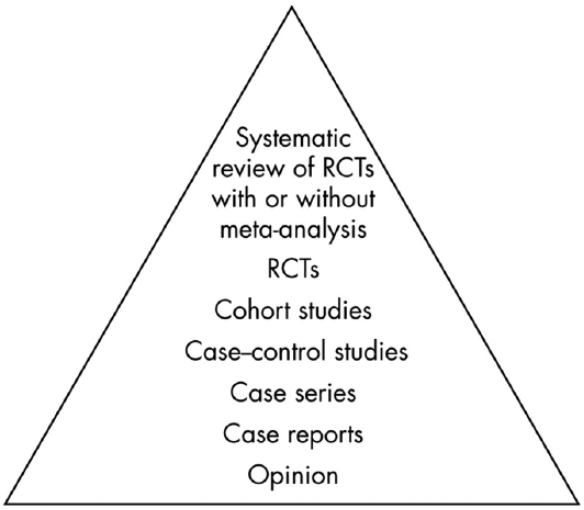

Figure 1. Hierarchy of research evidence.
RCT: randomised controlled trial

Steps

Choose appropriate guidelines for reporting a systematic review

It is important to follow accepted guidelines, as this is a requirement for publication in many journals. For meta-analyses of RCTs, this is the QUOROM Statement (QUality Of Reporting Of Meta-analyses). An updated version, called PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses), has extended coverage to systematic reviews.[2] Although, the PRISMA statement focuses on randomised trials, it can also be used for other designs, such as service evaluation, and consists of a checklist and flow diagram (Figure 3). For observational studies, there are recommendations for the reporting of MOOSE (Meta-analyses Of Observational Studies in Epidemiology).[3]

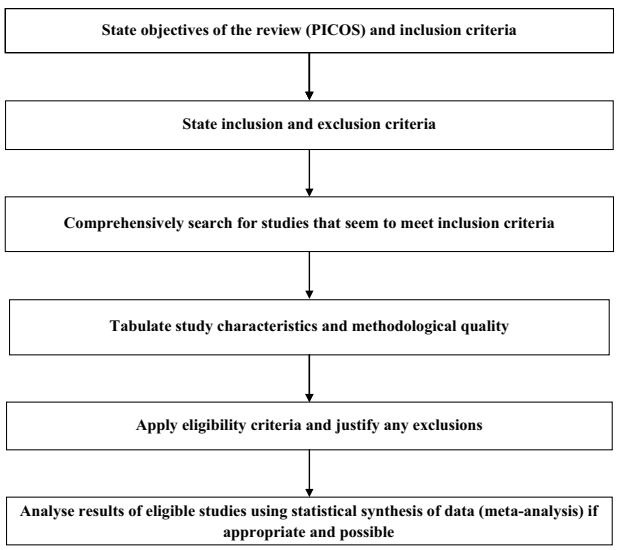

Figure 2. Methodology for a systematic review and meta-analysis (adapted from Akobeng[1]).
PICOS: Population, Intervention, Controls, Outcomes and Study design.

The research question

Like other research, there should be an explicit question that specifies the relevant Population, Intervention, Controls, Outcomes and Study design (PICOS). Avoid being too broad (e.g. anti-depressants for depression), as you may be overwhelmed with potential papers. By contrast, 'Deep Brain Stimulation for obsessive-compulsive disorder' was a discrete topic that formed part of a psychiatry trainee's scholarly project.[4]

Types of studies included in the review

The types of study will depend on the research question. For observational studies, these might include case control and cohort designs while experimental studies include controlled, quasi-experimental and randomised controlled trials. Results from randomised and non-randomised trials should not be combined.[5]

Search strategy

This needs to be explicit, include all relevant databases, and not be limited to English. The most commonly used databases are the *Cochrane controlled trials register, PubMed/Medline, Cinahl, Embase* and *PsycINFO*, supplemented by hand searches of selected journals and retrieved articles. Unpublished papers should also be sought, so as to minimise publication bias (see section below). Where possible, enlist a librarian's help.

Article selection and data extraction

These should be done independently, by at least two people, to avoid selection bias. Retrieved citations are first scrutinised to eliminate duplicate studies and to

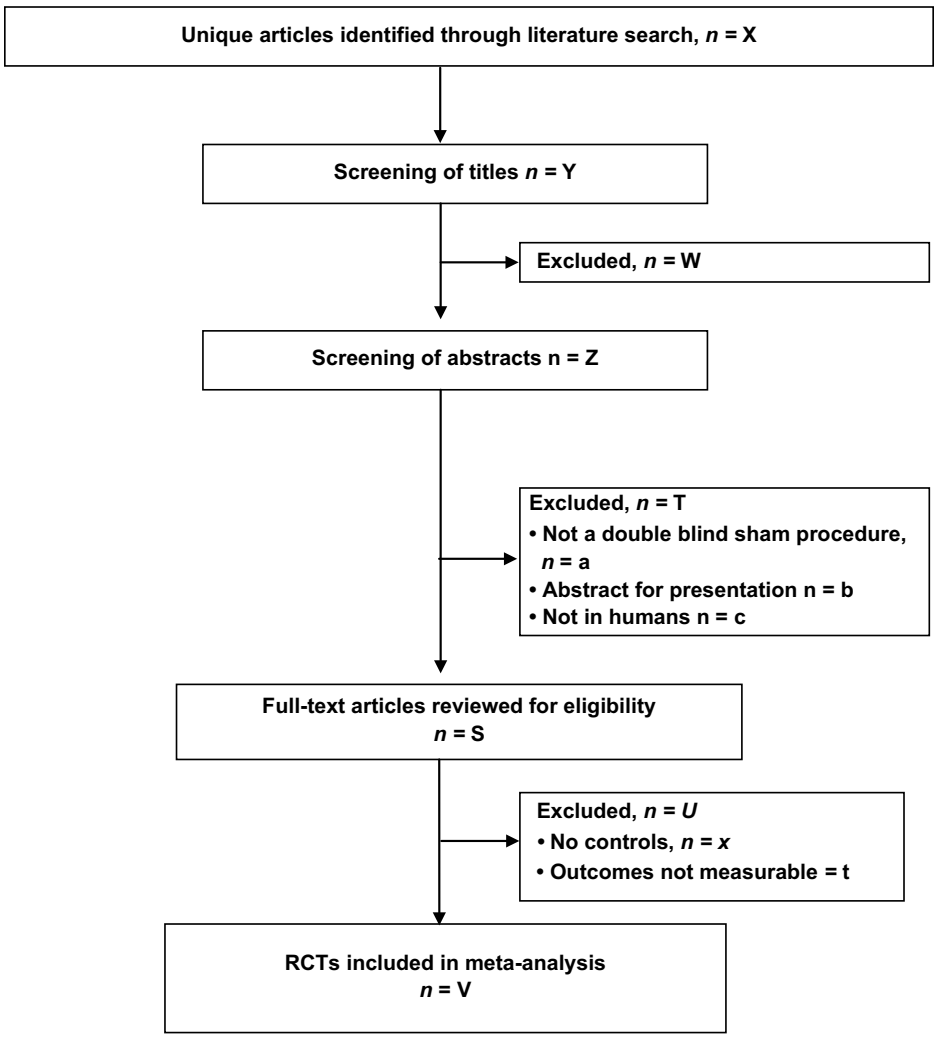

Figure 3. Sample PRISMA flow diagram.

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses; RCT: randomised controlled trial

see if there are any that can be excluded on the basis of the title alone. Next, abstracts are reviewed and a smaller number of studies selected where reviewing of the full-text paper is required. Finally, data from all included studies are entered into a table that includes the number of participants, their age and gender, setting, intervention and outcomes, as measured on standardised instruments.

For dichotomous variables, numerators and denominators are needed; while for continuous ones, it should be the mean and standard deviation or error. At this the stage, the quality of the studies should also be assessed. The following four criteria of the risk of bias assessment tool, developed by the Cochrane Collaboration, can be used to assess sources of bias in RCTs[5]:

- Adequate generation of allocation sequence;
- Concealment of allocation to conditions;
- Prevention of knowledge of the allocated intervention to assessors of outcome; and
- Dealing with incomplete outcome data.

The Newcastle-Ottawa Scale (NOS) is the equivalent for observational studies and it assesses quality in the selection of the study groups, their comparability and the ascertainment of outcome.[6]

The number of studies included and excluded at each stage should be recorded and presented in a flow diagram (Figure 3).

Analysis

Early attempts at quantitative analysis consisted of 'vote-counting' studies that showed benefit, harm or no difference; however, this method took no account of study

size, nor strength of the effect. It has now been replaced by meta-analysis, where data from several studies are combined, taking both into account. Meta-analyses increase statistical power, as determined by the *P* value; and improve precision, as shown by narrower confidence intervals (CI). They are especially valuable in situations where there are only small and possibly underpowered studies; and they can help establish the true efficacy of an intervention, where large studies may be impractical. However, there are times when meta-analyses should not be used, and it is inappropriate to define a systematic review as high quality, based solely on whether it contains a meta-analysis.

All meta-analyses are systematic reviews, but not all systematic reviews are meta-analyses. For instance, meta-analyses are inappropriate when studies or outcomes are clinically diverse, contain a mix of comparisons, are of poor quality, or in the presence of serious publication and reporting bias.

The ways in which an effect can be measured depend on the nature of the data. If the outcome is dichotomous (for example, disease versus no disease, remission versus no remission), then odds ratios (OR) or risk ratios (RR) are used. As a rough guide, the RR tends to be used in the meta-analyses of prospective studies, while the OR are used for case-controlled and cross-sectional studies.

If the outcome is continuous, and all the studies have used the same measure (e.g. Beck Depression Inventory scores (BDI)), mean differences are used. When studies report on the same phenomenon (e.g. depression), but use different scales (e.g. the BDI and Hamilton Rating Scale for Depression), the standardised mean difference is calculated.

All these are measures of relative effect, but sometimes RCTs will report measures of absolute effects, such as the risk difference or number-needed-to-treat. These should also be incorporated into a meta-analysis.

The typical graph for displaying the results of a meta-analysis is a 'forest plot'. These display the 95%CI of an estimate (e.g. odds or risk ratios), with the interval width indicating the precision of the estimate.

Clustered RCTs are a special case, as they randomise subjects at a programme or unit level, not individually; however, people within a programme or unit may resemble each other more than by chance (e.g. general practitioner clinics in affluent versus socially-deprived neighbourhoods), and so overestimate the potential benefits; therefore, the effect size of each study needs to be adjusted by using the intra-class correlation (ICC), a number from 0 to 0.99 derived from external databases.[7]

Of the available software, Review Manager (RevMan) is the most commonly used. It is a statistical package for Cochrane Collaboration systematic reviews and is downloadable free from the *Cochrane Collaboration*, along with an instruction manual.[8] It can handle all aspects of a Cochrane Review, including word processing, but unless you are writing a Cochrane Systematic Review, it is best to just use the analytic function. Figure 4 shows where these functions are located. References can be directly imported as text files from databases such as *Medline*, or from bibliographic software such as Endnote. Another software package is Comprehensive Meta-analysis (CMA).

Heterogeneity

When pooling the results of the individual studies in a meta-analysis, it is important to determine whether the studies were sufficiently similar for this to be reasonable. Differences or heterogeneity in studies can be:

- Clinical, in terms of participants, interventions, or outcomes;
- Methodological, in terms of trial design or quality, or;
- Statistical, resulting from clinical or methodological heterogeneity and where the observed effects are more different than that which can be expected by chance.

In the first instance, clinical judgement can help ensure that trials are sufficiently similar, in terms of participants, interventions, comparisons or outcomes. Heterogeneity should also be suspected if the results on the forest plot vary greatly in their direction. There are also statistical tests for heterogeneity, although their power may be insufficient to definitely exclude it. The Cochrane Collaboration favours the 'I^2 Statistic', which ranges from 0% to 100%, and where 0% means there is no heterogeneity.[4] Cuts-offs of 25%, 50% and 75% represent low, moderate and high heterogeneity, respectively. It is calculated using the Chi-square Statistic (Q), but unlike the Q Statistic alone, it does not depend on the number of papers in the meta-analysis; therefore, it has greater power to detect heterogeneity when there are relatively few studies.[5]

Other ways of dealing with statistical heterogeneity include further exploration through subgroup analysis (e.g. specific age groups) or excluding studies via a sensitivity analysis. Random effects models (see below) can incorporate heterogeneity to some extent. A final approach is meta-regression, where the effects of factors such as gender and age can be investigated as independent variables. This technique is available in CMA, but not in RevMan.

Fixed effects or random effects

There are two main techniques for summarising results. A 'fixed-effects model' assumes little study-to-study variability. By contrast, the 'random-effects model' assumes that effects may vary across studies. Fixed effects models are more likely to find significant

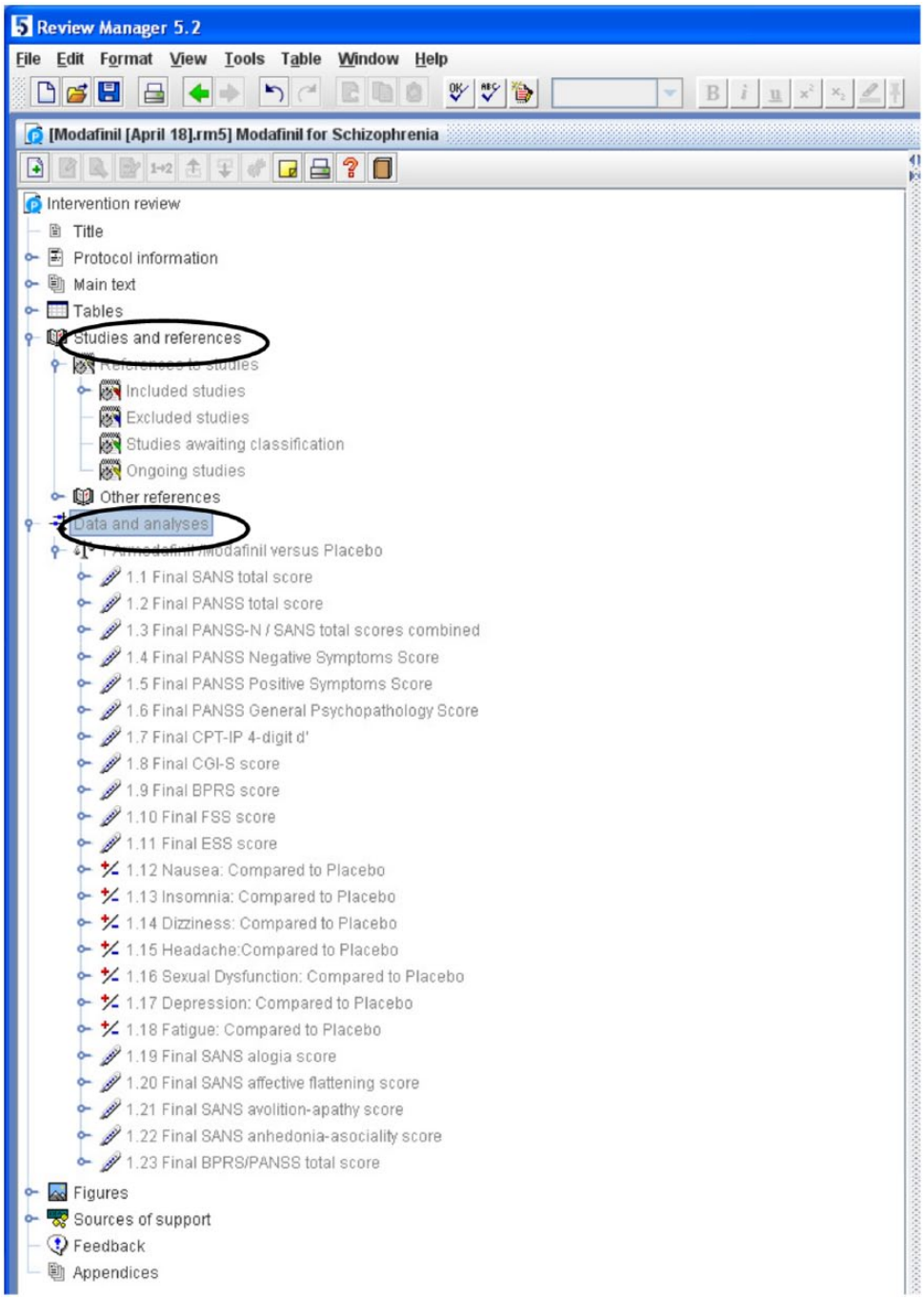

Figure 4. Screen shot of the RevMan software tree, showing the location of the bibliographic and analytical functions of the program.

differences, but are only appropriate when no clinical, methodological nor statistical heterogeneity is suspected. In this case, the random-effects model should always be used.

Publication bias

Publication bias occurs because studies with statistically significant results are more likely to be published and cited, and these are often preferentially published in English. Availability may also vary from being actively disseminated (e.g. pharmaceutical company-sponsored trials), easily available (e.g. open-access electronic journals), available in principle (e.g. limited-circulation print journals) or unavailable (such as unpublished data). These biases can be assessed with the 'Fail-safe *N*' or funnel plots. The fail-safe *N* statistic is the number of non-significant studies necessary to

reduce the odds ratio or affect size to a negligible value. It can be calculated using a free software package called Win-Pepi.[9] Funnel plots can be obtained using RevMan, with asymmetry suggesting publication bias. At least 10 studies are generally required for accurately estimating publication bias.

Cochrane versus non-Cochrane reviews

Cochrane systematic reviews have several advantages. Authors are guided through title registration, protocol submission and review preparation. Provided all the steps are followed, a paper in a publication with a good impact factor is practically guaranteed. Interest can be registered with one of the Cochrane groups, such as the schizophrenia, depression, anxiety or neurosis groups. Some will even do the literature search. Unfortunately, all of this comes at a price, which is the requirement to update the review every 2 years. This can become quite onerous, especially as the presentation of studies in the existing version may also have to be updated, in light of changing Cochrane requirements.

Conclusions

Systematic reviews and meta-analyses are not as daunting as they may appear, provided the scope of the review is sufficiently narrow and an appropriate supervisor is available.

Acknowledgements

This paper is based on workshops conducted by author SK at the Congress of the Royal Australian and New Zealand College of Psychiatrists (RANZCP), held on 12 – 15 May 2014 in Perth, Western Australia; and the annual conference of the Canadian Psychiatric Association, held on 13 – 15 October 2011 in Vancouver, Canada.

Disclosure

The authors report no conflict of interest. The authors alone are responsible for the content and writing of the paper.

References

1. Akobeng AK. Understanding systematic reviews and meta-analysis. *Arch Dis Child* 2005; 90: 845–848.
2. Moher D, Liberati A, Tetzlaff J, et al. The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009; 6(7): e1000097. doi:10.1371/journal.pmed.1000097
3. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *J Am Med Ass* 2000; 283: 2008–2012.
4. Kisely S, Hall K, Siskind D, et al. Deep Brain Stimulation for obsessive-compulsive disorder – A systematic review and meta-analysis. *Psychol Med* 2014; 44: 3533–3542.
5. Higgins JPT and Green S. *Cochrane handbook for systematic reviews of interventions, Version 5.1.0*. Oxford: The Cochrane Collaboration, 2011.
6. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses, 2000. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (2014, accessed 21 November 2014).
7. The Health Services Research Unit (HSRU). Research Tools, https://www.abdn.ac.uk/hsru/research/research-tools/ (2014, accessed 21 November 2014).
8. Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration. http://tech.cochrane.org/revman/about-revman-5 (2014, accessed 21 November 2014).
9. Abramson JH. WINPEPI updated: Computer programs for epidemiologists and their teaching potential. *Epidemiol Perspect Innovations* 2011; 8: 1.