REVIEW

BJC

British Journal of Cancer (2013) 108, 2299–2303 | doi: 10.1038/bjc.2013.185

Keywords: systematic review; meta-analysis; prognostic marker; diagnostic marker; sensitivity and specificity; hazard ratio

A step-by-step guide to the systematic review and meta-analysis of diagnostic and prognostic test accuracy evaluations

Z Liu[1,2], Z Yao[1], C Li[3], X Liu[1], H Chen[1] and C Gao*[,1]

[1]Anal-Colorectal Surgery Institute, 150th Central Hospital of PLA, Luoyang 471031, Henan, China; [2]Jinan Military General Hospital, Jinan 250031, China and [3]Division of head and neck surgery, Cancer Hospital of Sichuan, Sichuan 610000, China

In evidence-based medicine (EBM), systematic reviews and meta-analyses have been widely applied in biological and medical research. Moreover, the most popular application of meta-analyses in this field may be to examine diagnostic (sensitivity and specificity) and prognostic (hazard ratio (HR) and its variance, standard error (SE) or confidence interval (CI)) test accuracy. However, conducting such analyses requires not only a great deal of time but also an advanced professional knowledge of mathematics, statistics and computer science. Regarding the practical application of meta-analyses for diagnostic and prognostic markers, the majority of users are clinicians and biologists, most of whom are not skilled at mathematics and computer science in particular. Hence, it is necessary for these users to have a simplified version of a protocol to help them to quickly conduct meta-analyses of the accuracy of diagnostic and prognostic tests. The aim of this paper is to enable individuals who have never performed a meta-analysis to do so from scratch. The paper does not attempt to serve as a comprehensive theoretical guide but instead describes one rigorous way of conducting a meta-analysis for diagnostic and prognostic markers. Investigators who follow the outlined methods should be able to understand the basic ideas behind the steps taken, the meaning of the meta-analysis results obtained for diagnostic and prognostic markers and the scope of questions that can be answered with Systematic Reviews and Meta-Analyses (SRMA). The presented protocols have been successfully tested by clinicians without meta-analysis experience.

Systematic reviews of high-quality randomised controlled trials are crucial in evidence-based medicine (EBM), as they are particularly useful for overcoming the difficulties faced by clinicians when they wish to extract and analyse dates to guide their practice. Systematic reviews are of special value in aggregating and synthesising the findings of many separately conducted studies, sometimes with conflicting results (Clarke, 2007). The purpose of the preferred reporting Items for Systematic Reviews and Meta-Analyses (SRMA) guidelines is to aid researchers in the reporting of SRMA (Liberati *et al*, 2009; Moher *et al*, 2010). When a review makes an effort to comprehensively identify and trace all of the literature on a given topic (also referred to as a systematic literature review), meta-analysis is a particular statistical strategy for bringing together the results of several studies to produce a single estimate (Sackett *et al*, 2007).

Numerous reports and books have been published that describe SRMA, but several predominant problems still exist, as stated below.

(1) Dispersed and fragmentary documents that describe SRMA are much more numerous than systematic and comprehensive reports. Additionally, the materials available that introduce the theory of SRMA are much more numerous than those addressing methodology, and the theoretical and methodological papers relevant to SRMA are often independent of each other. Hence, obtaining comprehensive collections of the materials related to SRMA is time-consuming work.
(2) The articles and books related to SRMA focus primarily on techniques, but not practices, necessitating the need for a strong background in fields such as mathematics, statistics and

*Correspondence: Dr C Gao; E-mails: zhongyujohn@gmail.com and gaochunfang2010@163.com

Received 2 January 2013; revised 22 March 2013; accepted 24 March 2013; published online 21 May 2013

computer science. For non-technical readers and the majority of biologists and clinicians, these materials are daunting and will lead to a natural aversion to meta-analysis, thus hindering the wide application of SRMA.

(3) Biomarkers (especially disease markers) have been widely applied in biological and clinical analyses. The characteristic of a single marker are usually reported in many articles, and an emergent task is the integration of the effect sizes of markers (i.e., sensitivity and specificity of a diagnostic/screening marker or the hazard ratio (HR) and its variance, SE or CI as a prognostic/monitoring marker) using SRMA. Nevertheless, according to our investigations, few articles and software tools are available that fully elaborate a procedure to combine the effect sizes of markers.

To address the above problems, we collected various materials and compiled a protocol using non-technical language as much as possible to guide common audiences in a step-by-step manner to realise SRMA, aiming at effect sizes of biomarkers with zero barriers. Our goal is to make SRMA accessible to most audiences, including biologists, clinicians and novices. We believe that investigators who read our paper will benefit from our protocol.

In this article, given that our protocol only focuses on SRMA of biomarker test accuracy, the following descriptions are not strictly in accordance with the above-mentioned general eight-step method. Based on the characteristics of the effect size of the accuracy of diagnostic and prognostic tests, we provide a five-step workflow.

STEPS FOR SRMA OF BIOMARKERS

Search strategy. The goal of the literature search is to be sufficiently exhaustive to develop a comprehensive list of potentially relevant studies. The first step in a meta-analysis is to find all of the pertinent articles on your topic. Important sources of information for meta-analyses include MEDLINE (http://www.ncbi.nlm.nih.gov/pubmed), EMBASE (http://www.ncbi.nlm.nih.gov/pubmed), OvidSP (http://www.ovid.com/) and CancerLit (http://www.twu.ca/library/cancer-lit.htm). The Cochrane Collaboration Controlled Trials Register, established in 1993, is also an important source of studies for a meta-analysis. It includes all of the controlled trials in the MEDLINE and EMBASE as well as the results of manual searches conducted by Cochrane Collaboration volunteers of thousands of journals not indexed by MEDLINE or EMBASE. Before applying the literature search strategy, the basic information and search syntax should be mastered; key words related to your topic should be listed; and the 'associated words' for each key word must also be prepared (see Tamara Durec BSc(Pharm), 2013 and Literature Searching and Systematic Reviews (2013) for more details and Appendix 1 as an example in Supplementary Materials).

Inclusion and exclusion criteria. Biomarkers are commonly categorised into four types: screening, diagnostic, prognostic and monitoring (surveillance). The first two types of markers are assessed based on sensitivity and specificity in most cases, while prognostic and monitoring markers are estimated based on the HR and its variance or standard error (SE) because they are time-to-event markers. Once the author of a meta-analysis has assembled a large number of studies, it is important to select the right ones. Which studies are included or excluded depends on various factors, such as whether or not there is sufficient information in a study to conduct an analysis, in addition to the study design, dosage used in the study, sample size, patient age, and even the year of the study. The following general criteria are provided only for reference:

(1) The study should be an original report (i.e., letters, editorials, case reports, tutorials and reviews are excluded), and both English and non-English studies should be included in case of a publication bias.
(2) The study should assess the ability of one or more markers to detect the presence of a particular disease.
(3) The study should provide sufficient data to allow estimation of a marker's accuracy, for a diagnostic marker, the study must directly or indirectly provide at least four values, which are the following: the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), to (re)construct a two-by-two table. (See the diagnostic marker sheet of Supplementary Table 1 for other relevant information requirements.) For a prognostic marker, the study must directly or indirectly provide at least two values, the HR, and its variance and/or SE and/or confidence interval (CI). (See the prognostic marker sheet of Supplementary Table 1 for other relevant information requirements.)
(4) Combination marker and review articles are excluded.
(5) If multiple papers are published based on the same or overlapping data sets, then only the paper with the largest number of specimens, the most detailed results and the longest follow-up time is included.

A minimum of two reviewers perform a first-stage screening of titles and abstracts based on the research question and the study design, population, intervention and outcome to be studied. Based on the initial screening, selected full-text articles are obtained for the second-stage screening. Including two reviewers minimises the introduction of bias by either reviewer. Any study identified by either reviewer should be included. Using the full text, a second-stage screening is performed by at least two reviewers. The studies selected are then submitted for data extraction.

Data extraction. Once an appropriate group of studies has been identified, the relevant data from candidate studies must be correctly extracted. To minimise errors, the following conventions should be considered: (1) all reviewers must be trained under a consensus standard and then practice using several articles for 'calibration'; (2) a consensus form or database that constrains entries to the expected range should be determined in advance; (3) at least two independent reviewers should check and extract data from a given article, and if the extracted data are not same, conflicts are resolved by reaching a consensus; and (4) to prevent bias creeping into a meta-analysis, the reviewers should not be biased in favor of (or against) well-known researchers or prominent journals as far as possible.

For studies related to biomarkers, the reviewers should also pay attention to the following matters in addition to the preceding items. Among the four types of markers, screening and diagnostic markers mainly focus on sensitivity and specificity, while prognostic and monitoring markers are usually focused on the HR and its variance, SE or CI. Consequently, we will classify the markers into two different categories and describe how to abstract relevant information: (1) if more than one marker is used in a given study, then the relevant data for each eligible marker must be individually extracted; (2) if one marker has multiple functions (i.e., one marker for one disease is used for screening, diagnosis, prognosis and/or monitoring), then the data sets corresponding to multiple functions must be extracted separately; and (3) if there are multiple markers and diseases addressed in one study, then only the relevant data from the marker(s) corresponding to each disease of interest for the author(s) should be extracted.

For screening or diagnostic markers. For diagnostic (or screening) markers, the data abstraction phase involves an assessment of study quality. Whiting *et al* (2003) proposed a set of criteria for

Quality Assessment of Diagnostic Accuracy Studies (QUADAS) that applies well to diagnostic marker studies (Whiting *et al*, 2003). Data extraction (see Appendix 2 in Supplementary Materials for details) with QUADAS assessment is often completed at the same time.

For prognostic or monitoring markers. In contrast, data extraction and conversion for prognostic (or monitoring) markers are much more complex than for diagnostic markers because prognostic markers provide time-to-event data, which indicate the distinctions between the two groups of studies, and time to event needs to be scrutinised very carefully since the data may not only be right censored (patient was not followed until the event), but also left censored (patients were not all followed starting from a comparable point). Meta-analyses of this type of marker often require one of two kinds of data, that is, the log of the HR (namely, $\log_e$(HR)) and its variance or SE or the HR and its CI. For major prognostic marker studies, the two kinds of data cannot be extracted directly. Parmar *et al* (1998) presented a series of simple methods to extract relevant data from publications with the aim of performing a meta-analysis of survival-type data. The methods focus on approaches for extracting these data from publications and are illustrated throughout this publication with real examples. Riley *et al* (2003) summarised 11 methods (see Appendix 3 in Supplementary Materials for details) that are available for directly or indirectly estimating these data and the approximate normal $\log_e$(HR) distribution for large samples. In addition, Tierney *et al* (2007) provided step-by-step guidance for how to calculate an HR and the associated statistics for individual trials, according to the information presented in the trial report.

Statistical methods. When studies used a similar design, we often combine the information they provide to increase precision and to investigate consistencies and discrepancies between the results. There has been great growth in this kind of analysis in several fields in recent years, particularly in medicine. In medicine, such studies usually involve controlled therapeutic trials. We apply the same principles in any scientific area, such as epidemiology, psychology or educational research. The essence of meta-analysis is obtaining a single estimate of the effect size from each similar study. There are many issues and controversies regarding meta-analysis data. First, we have to define two important terms, homogeneity and heterogeneity, to describe the degree of between-study variability in a group of studies. Fixed-effect models consider only within-study variability. The assumption is that studies use identical methods, patients and measurements; that they should produce identical results; and that any differences are only due to within-study variation. Random-effect models consider both the between-study and within-study variability. It is assumed that studies provide a random sample from the universe of all possible studies. If the studies are heterogeneous, then a random-effect model is applied for meta-analysis of the effect size in a group of studies; otherwise, a fixed-effect model is selected (see Appendix 4 in Supplementary Materials for detailed interpretations). A meta-analysis will customarily include a forest plot, in which the results from each study are displayed as a square and a horizontal line, representing the intervention effect estimated together with its CI. The area of the square reflects the weight that the study contributes to the meta-analysis. The combined-effects estimate and its CI are represented by a diamond. Biomarkers generally include screening, diagnostic, prognostic and monitoring markers. The first two types of markers correspond to diagnostic tests, and the last two provide time-to-event data. In meta-analyses of the two kinds of tests, there are significant differences in terms of both the combined objects and methods. Hence, descriptions of the two kinds of meta-analyses are provided below.

Analysis of diagnostic (or screening) test accuracy. The meta-analysis of diagnostic test accuracy represents an area of growing interest. These analyses often consist of three steps: assessment of study quality, creation of forest plots for sensitivity and specificity for each study, and summarisation of estimates of sensitivity and specificity using two types of models.

Quality assessment for diagnostic articles: Quality assessment is as important in systematic reviews of diagnostic accuracy studies as it is in any other type of review, and the methodological quality of each study was assessed as recommended by the Cochrane Diagnostic Test Accuracy Working Group. These recommendations were adapted from the QUADAS guidelines (Whiting *et al*, 2003; Macaskill *et al*, 2010). All of the criteria were classified as Yes, No or Unclear based on information available in this publication (see Appendix 5 in Supplementary Materials). The studies were judged according to the data used for the meta-analysis, which may not include all of the data available in the publication.

Meta-analysis of the accuracy of diagnostic tests: Meta-analyses of diagnostic test accuracy present many challengers: (1) even in the simplest case, a minimum of two summary statistics (sensitivity and specificity) must be addressed simultaneously; (2) meta-analysis methods allow studies to be combined that have applied tests at different thresholds; and (3) random-effect methods are recommended when data are heterogeneous (this is the rule for diagnostic studies). Therefore, in a meta-analysis of diagnostic accuracy, two analysis steps must be completed: (1) forest plots for pooling the sensitivity and specificity of all of the selected studies are first created; and (2) two statistical methods to calculate summary estimates of sensitivity and specificity are proposed to account for the correlation between sensitivity and specificity across studies caused by the relationship between sensitivity and specificity within each study (Moses *et al*, 1993). The two statistical methods, that is, the hierarchical summary receiver operating characteristic (HSROC) model (Rutter and Gatsonis, 2001) and bivariate model (Reitsma *et al*, 2005), are statistically rigorous (Appendix 6 in Supplementary Materials introduces two ways to perform a meta-analysis of the accuracy of diagnostic tests).

Meta-analysis of the accuracy of prognostic (or monitoring) tests. For prognostic (monitoring) markers, as described above, there are two types of extracted data: the HR and its CI (lower limit and upper limit); and $\log_e$(HR) and its variance (var[$\log_e$(HR)]) or standard error (SE[$\log_e$(HR)] = the reciprocal of the square root of var[$\log_e$(HR)]). Among the existing meta-analysis software tools, RevMan 5.1 (Review Manager, 2011 and MetaDisc 1.4 (Zamora *et al*, 2006) cannot be used to implement a meta-analysis of the accuracy of prognostic tests. The mada package in the R language can be used to perform meta-analysis of prognostic test accuracy, but the R language requires complete user-entry of codes. In contrast, the mode of operation of STATA involves an interface of window plus commands that can be used by common audiences, making the performance of a meta-analysis using STATA (see Appendix 7 in Supplementary Materials for STATA installation and 14 STATA meta-analysis commands) much easier than using the R language. Next, we will use non-technical language to interpret how to perform a meta-analysis of the accuracy of prognostic tests.

Meta-analysis of prognostic test accuracy: Meta-analysis is a two-stage process involving the estimation of an appropriate summary statistic for each of a set of studies, followed by the calculation of a weighted average of these statistics across the studies (Deeks *et al*, 2008).

The summary statistics from each study can be combined using a variety of meta-analytical methods, which are classified as fixed-effect models in which studies are weighted according to the amount of information they contain, or random-effects models, which incorporate an estimate of between-study variation (heterogeneity) in the weighting. A meta-analysis will customarily include a forest plot, where the results from each study are displayed as a square and a horizontal line, representing the intervention effect estimate together with its CI. The area of the square reflects the weight that the study contributes to the meta-analysis. The combined-effect estimate and its CI are represented by a diamond. Here, we present updates to the *metan* command in STATA to perform a meta-analysis of prognostic test accuracy. *metan* provides methods for the meta-analysis of studies with two groups, and either fixed-effect or random-effect models can be fitted (Fleiss, 1993). The following is the syntax for metan: *metan [varlist] [option]*. In Supplementary Materials document, we used two different data types as examples (see examples 1 and 2 in Supplementary Materials for the detailed operation flows) to present the *metan* command for the meta-analysis of prognostic test accuracy.

Publication bias regarding prognostic test accuracy: Publication bias is the phenomenon of studies with uninteresting or unfavorable results being less likely to be published than those with more favorable results (Rothstein *et al*, 2005). If a publication bias exists, then the published literature is a biased sample of all studies on a topic, and any meta-analysis based on it will be similarly biased. Funnel plots are commonly used to investigate publication and related biases in meta-analyses (Sterne *et al*, 2005). **metabias** performs the Begg and Mazumdar (1994)) adjusted rank correlation test for publication bias as well as the Egger *et al* (1997) regression asymmetry test for publication bias. As options, it provides a funnel graph of the data or the regression asymmetry plot. The Begg adjusted rank correlation test is more popular in common applications for publication bias analysis (see examples 3 and 4 in Supplementary Materials).

Non-parametric trim and fill analysis of publication bias. Meta-analysis is a popular technique for numerically synthesising information from published studies. One of the many concerns that must be addressed when performing a meta-analysis is whether selective publication of studies could lead to a bias in estimating the overall meta-analytical effect and in the inferences derived from the analysis. If a publication bias appears to exist, then it is desirable to consider what the unbiased data set might look like and then to re-estimate the overall meta-analytical effect after any apparently 'missing' studies are included. Duval and Tweedie's 'non-parametric 'trim and fill' method' is designed to meet these objectives (Duval and Tweedie, 2000). The command metatrim is used to implement the Duval and Tweedie non-parametric 'trim and fill' method (see examples 5 and 6 in Supplementary Materials).

Cumulative meta-analysis of prognostic test accuracy: In a cumulative meta-analysis (Rothstein *et al*, 2005), the pooled estimate of the treatment effect is updated each time the results of a new study are published. This makes it possible to track the accumulation of evidence related to the effect of a particular treatment. The command metacum performs a cumulative meta-analysis (using fixed- or random-effect models), and optionally, the results can be graphed. A user supplies the preceding two types of data on prognostic test accuracy. The full metacum command is very similar to the metan command. The detailed commands are as follows:

metacum lnhr lnll lnul, eform label (namevar = studyid) title ("random-effect model") boxsca (0.9) random effect (Hazard Ratio)

(The output and forest plot are omitted)

metacum logehr selogehr, eform effect (Hazard Ratio) title ("Fixed-effect meta-analysis") boxsca(0.9) label (namevar = paperno)

(The output and forest plot are omitted).

Subgroup for prognostic test accuracy: Dividing results between different types of patients and outcomes requires cautious interpretation. If these analyses are to be conducted, then more subgroup analyses must be performed. It is often more reliable to assume that the overall result is as good an estimate (if not a better one) for a particular group of patients than that obtained by examining those patients within the meta-analysis. In prognostic test data, it is very common for the data to be classified into disease-free survival (DFS) and overall survival (OS) subgroups, as in the data in Example data 1. In fact, performing both subgroup analyses in meta-analysis in STATA is very simple, and a major addition to metan is the ability to perform stratified or subgroup analyses. Subgroup analyses in meta-analysis may be used to investigate the possibility that treatment effects vary between subgroups. Subgrouping in meta-analysis can be completed by adding one option: *by (grouping variable name)*, to all meta-analysis commands for diagnostic or prognostic test accuracy (all examples are omitted).

CONCLUSION

In EBM, SRMA have been widely applied in biological and medical research. Moreover, the most popular application of meta-analysis in this field may be to assess diagnostic (sensitivity and specificity) and prognostic (the HR and its precision) test accuracy. With the growth of clinical renal studies, an increasing number of these types of summary publications will certainly become available to nephrologists, researchers, administrators and policy makers who seek to keep abreast of recent developments. To maximise the advantages of these studies, it is necessary for these individuals to have a simplified version of a protocol to aid them in rapidly conducting meta-analyses of the accuracy of diagnostic and prognostic tests. In this article, we first presented a simplified and practical protocol to guide non-professional academicians and clinicians to perform systematic reviews of diagnostic and prognostic accuracies in a step-by-step manner, and we confirmed that once an individual studies our article, even a novice, they are soon able to accomplish complex systematic reviews and meta-analyses. The protocols have been successfully tested by clinicians without meta-analysis experience.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for constructive comments on the manuscript. Funding for this work was provided by China Postdoctoral Science Foundation (201150M1569 and 2012T50893 to ZL).

AUTHOR CONTRIBUTIONS

ZL contributed to the design, preparation and editing of the document. CG was responsible for the final review and approval for submission.

REFERENCES

Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**: 1088–1101.

Clarke M (2007) The Cochrane Collaboration and systematic reviews. *Br J Surg* **94**(4): 391–392.

Deeks JJ, Altman DG, Bradburn MJ (2008) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In *Systematic Reviews in Health Care: Meta-analysis in Context*. 2nd edn Egger M, Smith GD, Altman DG (eds) BMJ Publishing Group: London, UK doi: 10.1002/9780470693926.ch15).

Duval S, Tweedie R (2000) A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *J Am Stat Assoc* **95**: 89–98.

Egger M, Smith GD, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *Br Med J* **315**: 629–634.

Fleiss JL (1993) The statistical basis of meta-analysis. *Stat Methods Med Res* **2**: 121–145.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* **21**: 339 b2700.

Literature Searching and Systematic Reviews (2013) Available from www.rds-eoe.nihr.ac.uk/activity/docs/infosheet-5.doc (accessed 12 March 2013).

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y (2010) Chapter 10: analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C (eds), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration: Oxford, UK 1–47. Available from http://srdta.cochrane.org/.

Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA Group (2010) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* **8**(5): 336–341.

Moses LE, Shapiro D, Littenberg B (1993) Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* **12**(14): 1293–1316.

Parmar MK, Torri V, Stewart L (1998) Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* **17**(24): 2815–2834.

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* **58**(10): 982–990.

Review Manager (RevMan) [computer software] (2011) *Version 5.1*. Cochrane Collaboration: Copenhagen, Denmark.

Riley RD, Burchill SA, Abrams KR, Heney D, Lambert PC, Jones DR, Sutton AJ, Young B, Wailoo AJ, Lewis IJ (2003) A systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma. *Health Technol Assess* **7**(5): 1–162.

Rothstein HR, Sutton AJ, Borenstein M (2005) Publication bias in meta-analysis: prevention. *Assessment and Adjustments* Chichester, UK, Wiley.

Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* **20**(19): 2865–2884.

Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (2007) Evidence based medicine: what it is and what it isn't.2006. *Clin Orthop Relat Res* **455**: 3–5.

Sterne JA, Becker BJ, Egger M (2005) The funnel plot. In *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, Rothstein HR, Sutton AJ, Borenstein M (eds), pp 75–98. Wiley: Chichester, UK.

Tamara Durec BSc(Pharm) (2013) Sarah Curtis—Online EBM Tutorial. Literature Searching For Systematic Reviews. Available from www.columbia.edu/∼mvp19/RMC/M2/Files/LitSearch.doc (accessed 12 March 2013).

Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR (2007) Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* **8**: 16.

Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* **3**: 25.

Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A (2006) Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* **6**: 31.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)