

be in keeping with the legal interpretation of good practice. Under the principles of the NHS plan, compassionate use could also be interpreted as a flexible, more responsive approach, but it could equally be viewed as a way around the evidence.

Individuals who persistently campaign for access to treatment are likely to succeed. These people are by nature assertive and articulate and are unlikely to be "average" consumers with a balanced view of the evidence. Consumers think as individuals and may take a different view from the proponents of evidence based medicine about the common good and best use of public funds. New structures are emerging from the NHS plan that purport to offer a modern partnership between the public and the NHS, such as independent local advisory forums.⁴ These forums will be a challenge because consumers are still poorly placed to understand differences of opinion between doctors on the evidence of treatments that have marginal benefit, particularly if consensus is used to value a treatment. It will take a lot of effort to find and equip a lay panel to

make informed contributions every time an urgent demand for a disputed treatment arises. Recognition of the tension between consumers and their advocates, the population need, and clinicians must lead to a more reflective public discussion about the trade-offs in achieving modernisation of the health service.

I thank David Strachan of St George's Hospital Medical School and Nick Mays for advice.

Competing interests: None declared.

- 1 Melzer D. New drug treatment for Alzheimer's disease: lessons from healthcare policy. *BMJ* 1998;316:762-4.
- 2 National Institute for Clinical Excellence. *Donepezil, rivastigmine, and galantamine for the treatment of Alzheimer's disease*. London: NICE, 2001. (Technology appraisal No 19.)
- 3 Standing Medical Advisory Committee. *The use of donepezil for Alzheimer's disease*. London: Department of Health, 1998.
- 4 Department of Health. *The NHS plan*. London: Stationery Office, 2000.
- 5 Birks JS, Melzer D. Donepezil for mild and moderate Alzheimer's disease. *Cochrane Database Syst Rev* 2000;(4):CD001190.
- 6 Donepezil update. *Drug Ther Bull* 1998;36:60-1.
- 7 Rosen R, Mays N. The impact of the UK NHS purchaser-provider split on the "rational" introduction of new medical technologies. *Health Policy* 1998;43:103-23.

(Accepted 4 June 2001)

Systematic reviews in health care

Systematic reviews of evaluations of prognostic variables

Douglas G Altman

This is the last in a series of four articles

Imperial Cancer Research Fund Medical Statistics Group, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford OX3 7LF
Douglas G Altman
director
Series editor:
Matthias Egger

BMJ 2001;323:224-8

Prognostic studies include clinical studies of variables predictive of future events as well as epidemiological studies of aetiological risk factors. As multiple similar studies accumulate it becomes increasingly important to identify and evaluate all of the relevant studies to develop a more reliable overall assessment. For prognostic studies this is not straightforward.

Box 1 summarises the clinical importance of information on prognostic factors. Many of the issues discussed are also relevant to aetiological studies, especially cohort ones. Some features of prognostic studies lead to particular difficulties for the systematic reviewer. Firstly, in most clinical prognostic studies the outcome of primary interest is the time to an event, often death. Meta-analysis of such studies is rather more difficult than that for binary data or continuous measurements. Secondly, in many contexts the prognostic variable of interest is often one of several prognostic variables. When examining a variable of interest researchers should consider other prognostic variables with which it might be correlated. Thirdly, many prognostic factors are continuous variables, for which researchers use a wide variety of methods of analysis.

The emphasis in this paper is on clinical studies to examine the variation in prognosis in relation to a single putative prognostic variable of interest (also called a prognostic marker or factor). A more detailed discussion can be found elsewhere.²

Identifying relevant publications

It is probably more difficult to identify all prognostic studies by searching the literature than it is for randomised trials, which itself is problematic. As yet

Summary points

Systematic reviews are applicable to all types of research design, and studies of prognostic variables are an important additional area where appropriate methodology should be applied

Prognostic variables should be evaluated in a representative sample of patients assembled at a common point in the course of their disease—ideally they should all have received the same medical treatment or been in a randomised trial

When examined critically, a high proportion of prognostic studies are found to be methodologically poor

Meta-analysis of published data is hampered by difficulties in extraction of data and variation in the characteristics of the study and patients

The poor quality of the published literature is a strong argument in favour of systematic reviews but also an argument against formal meta-analysis

Meta-analysis of prognostic studies using individual data from patients may overcome many of these difficulties

there is no widely acknowledged optimal strategy for searching the literature for prognostic studies, but

Box 1—Purpose of studies of prognostic factors (adapted from Altman and Lyman¹)

- To guide clinical decision making, including treatment selection and patient counselling
- To improve understanding of the disease process
- To improve the design and analysis of clinical trials (for example, risk stratification)
- To assist in comparing outcome between treatment groups in non-randomised studies by allowing adjustment for case mix
- To define risk groups based on prognosis
- To predict disease outcome more accurately or parsimoniously

search strategies have been developed with either good sensitivity or good specificity (box 2).

Epidemiological studies are more prone to publication bias than randomised trials.⁴ It is probable that studies showing a strong (often statistically significant) prognostic ability are more likely to be published. Publication bias has recently been shown in studies of Barrett's oesophagus as a risk factor for cancer.⁵

Assessing methodological quality—design**Generic criteria**

There are no widely agreed quality criteria for assessing prognostic studies. As yet there is little empirical evidence to support the importance of particular study features affecting the reliability of findings, including the avoidance of bias. As a consequence, systematic reviewers tend either to ignore the issue or to devise their own criteria. Unfortunately the number of different criteria and scales is likely to continue to increase and cause confusion, as has happened for randomised trials and systematic reviews.⁶⁻⁸ Nevertheless, theoretical considerations and common sense point to several methodological aspects that are likely to be important. The table shows a list of those relating to internal validity, which draws on previous suggestions.⁹⁻¹³

A reliable prognostic study requires a well defined cohort of patients at the same stage of their disease. Some authors suggest that the sample should be an "inception" cohort of patients early in the course of the disease (perhaps at diagnosis).⁹ Whereas homogeneity is often desirable, heterogeneous cohorts can be stratified in the analysis. Also, not all prognostic studies

Box 2—Effective strategies for searching Medline for prognostic studies³**Best single term**

"Explode cohort studies" (MeSH)

Best complex search strategy with the highest sensitivity

"incidence" (MeSH)
 or "explode mortality" (MeSH)
 or "follow-up studies" (MeSH)
 or "mortality" (subheading)
 or "prognos*" (text word)
 or "predict*" (text word)
 or "course" (text word)



MARK OLDROYD

relate to patients with overt disease. An example is a study of prognostic factors in a cohort of asymptomatic patients infected with HIV.¹⁴

Both case-control and cross sectional studies may be used to examine risk factors, but these designs are much weaker. Case-control designs have been shown to yield optimistic results for evaluations of diagnostic tests, a result that is likely to be relevant to prognostic studies.¹⁵ In cross sectional studies it may be difficult to determine whether the exposure or outcome came first—for example, in studies examining the association between the use of oral contraceptives and HIV infection.

Windeler observed that summaries of prognosis are not meaningful unless associated with a particular strategy for treatment and suggested that the greatest importance of prognostic studies is to aid decisions about treatment.¹⁶ Most published checklists do not, however, consider the issue of subsequent treatment. If the treatment received varies in relation to prognostic variables then the study cannot deliver an unbiased and meaningful assessment of prognostic ability unless the different treatments are equally effective (in which case why vary the treatment?). Such variation in treatment may be quite common once there is evidence (usually non-systematic) that a variable is prognostic. Ideally, therefore, prognostic variables should be evaluated either in a cohort of patients treated the same way or in a randomised trial.^{12 17}

Criteria specific to studies

The inclusion of context specific as well as generic aspects of methodological quality is sometimes sensible. For example, a review of prognosis of idiopathic membranous nephropathy included two questions on the nature of the end points, reflecting particular problems in a discipline where many studies used ill defined surrogate end points.¹³

In addition to internal validity some checklists consider aspects of external validity and clinical usefulness of studies. Notably, Laupacis et al included five questions relating to the clinical usefulness of a study.¹¹ Furthermore, some checklists reasonably include items relating to the clinical area of the review. For example, in their review of the association between maternal HIV infection and perinatal outcome, Brocklehurst

Framework for assessing internal validity of articles dealing with prognosis*

Study feature	Qualities sought
Sample of patients	Inclusion criteria defined, sample selection explained, adequate description of diagnostic criteria, clinical and demographic characteristics fully described, representative, assembled at common (usually early) point in course of disease, complete (all eligible patients included)
Follow up of patients	Sufficiently long
Outcome	Objective, unbiased (for example, assessment blinded to prognostic information), fully defined, appropriate, known for all or high proportion of patients
Prognostic variable	Fully defined, including details of measurement methods if relevant, precisely measured, available for all or high proportion of patients
Analysis	Continuous predictor variable analysed appropriately, statistical adjustment for all important prognostic factors
Treatment subsequent to inclusion in cohort	Fully described, treatment standardised or randomised

*Partly based on references 9-13.

and French considered whether there was an adequate description of the maternal stage of disease.¹⁸

Assessing methodological quality—analysis

The table includes two items relating to difficult aspects of data analysis. It is important to adjust for other prognostic variables because patients with different values of the covariate of primary interest are likely to differ with respect to other prognostic variables. This procedure is often referred to as control of confounding. In contexts where much is known about prognosis, such as many cancers, it is important to know whether the variable of primary interest (such as a new tumour marker) offers prognostic value over and above that which can be achieved with previously identified prognostic variables. It follows that prognostic studies generally require some sort of multiple regression analysis. Comparison of models with and without the variable of interest provides an estimate of its independent effect and a test of whether it contains additional prognostic information.

Two problems for the systematic reviewer are that different researchers use different statistical approaches to adjustment and adjust for different selections of variables. One way around the second of these problems is to use unadjusted analyses. This approach is sensible in systematic reviews of randomised controlled trials, but in prognostic studies it replaces one problem with a worse one. Whereas the least adjusted estimate “provides the maximum opportunity for comparison of consistent estimates across studies,”¹⁹ unadjusted analyses will generally be biased.

Many prognostic variables are continuous measurements, including many biochemical and physiological measurements, tumour markers, and levels of environmental exposure. If such a variable were prognostic the risk of an event would usually be expected to increase or decrease systematically as the level increases. Keeping variables continuous can greatly simplify any subsequent meta-analysis, but most researchers prefer to categorise patients into high risk and low risk groups based on some cut-off point. This type of analysis discards potentially important quantitative information and considerably reduces the power to detect a real association with outcome.^{20 21} If a cut-off point is used it should not be determined by a data dependent process (such as exploring all cut-off points to find the one that minimises the P value).²²

The extraction of data is an additional problem. Some authors do not present a numerical summary of the prognostic strength of a variable, such as a hazard ratio, unless the analysis showed that the effect of that variable was significant. Also, when numerical results are given they may vary in format—for example, survival proportions may be given for different time points.

Meta-analysis

Box 3 summarises the particular difficulties for the systematic reviewer of prognostic studies. Two major concerns are the quality of the primary studies and the possibility of publication bias. Because of the likelihood of serious methodological difficulties, in general it is difficult to carry out a sensible meta-analysis without access to the data of individual patients.^{2 23} Many authors have concluded that a set of studies was too diverse or too poor (or both) to allow a meaningful meta-analysis. Box 4 summarises a systematic review of prognosis in elbow disorders, which reached such a conclusion. In a systematic review of the possible relation between hormonal contraception and risk of transmission of HIV, Stephenson concluded that a meta-analysis was unwise.²⁴ By contrast, Wang et al performed such a meta-analysis on a similar set of

Box 3—Problems with systematic reviews of prognostic studies from publications

- Difficulty of identifying all studies
- Negative (non-significant) results may not be reported (publication bias)
- Inadequate reporting of methods
- Variation in study design
- Most studies are retrospective
- Variation in inclusion criteria
- Lack of recognised criteria for quality assessment
- Different assays or measurement techniques
- Variation in methods of analysis
- Differing methods of handling of continuous variables (some dependent on data)
- Different statistical methods of adjustment
- Adjustment for different sets of variables
- Inadequate reporting of quantitative information on outcome
- Variation in presentation of results (for example, survival at different time points)

Box 4—Case study: prognosis in elbow disorders

Hudak et al carried out a systematic review of the evidence regarding prognostic factors that affect the duration of elbow pain and outcomes.¹² Selected papers were subjected to a detailed quality assessment by using a scheme adapted from other publications. Each paper was assessed on six dimensions: case definition, patient selection, follow up (completeness and duration), outcome, information about prognostic factors, and analysis. Each dimension was scored from 0 to 2 or 3. The authors' prespecified minimum requirements for studies providing strong evidence (number of studies out of 40 meeting the criteria in brackets) were:

- Provided an operational definition of cases (15)
- Included an inception cohort (defined in relation to onset of symptoms) or a survival cohort that included a subset of patients in whom duration of symptoms was less than 4 months (5)
- Showed follow up of more than 80% of cases for at least 1 year (8)
- Used a blinded and potentially replicable outcome measure appropriate to the research question (20)
- Used adequate measurement and reporting of potential prognostic factors (36)
- Provided crude proportions for at least one of response, recovery, and recurrence (34)

Papers were identified from a comprehensive literature search of multiple databases. The authors included the search strategy they used.

Of the 40 eligible studies assessed using the above criteria, none provided "strong evidence" and just four provided "moderate evidence," none of which followed patients for more than one year. The authors note that several studies with excellent follow up were not based on inception cohorts. Only three of the 40 studies had used a statistical method to derive results adjusted for other factors.

Among the four providing evidence of a "moderate level" there was variation in study design (one case series, three randomised trials), patient selection, interventions, and length of follow up. As a consequence meta-analysis was not attempted. The authors made several suggestions for the methodological requirements for future studies.

studies, arguing that this enabled the quantitative investigation of the impact of various features of the study.²⁵

Even when a set of published studies is of high quality there are many potential barriers to a successful meta-analysis. In essence it is desirable to compare the outcome for groups with different values of the prognostic variable. In principle it should be relatively easy to combine data from studies that have produced compatible estimates of effect with standard errors. In practice, the lack of comparable information from all studies is likely. In particular, the prognostic variable is likely to have been handled in various ways. In the simplest case, researchers may all have dichotomised but used different cut-off points. A meta-analysis is possible comparing "high" and "low" values, using whatever definition was used in the primary studies. Interpretation is difficult, because patients with the same values would be high in some studies and low in others. (This analysis will be biased if any studies used a cut-off point derived by the minimum P value method.) Studies may use different numbers of categories,²⁶ and

some may have categorised whereas others did not. Estimates derived from categorised and ungrouped analyses are not comparable.

Outcome is occurrence of event, regardless of time

As noted, in general it is necessary to allow for other potentially confounding variables in a meta-analysis. When time to an event is not relevant, logistic regression for both binary and continuous prognostic variables is used to derive an odds ratio after adjustment for other prognostic or potentially confounding variables. The adjusted odds ratio and confidence interval can be obtained from an estimated log odds ratio with its standard error. For a binary prognostic variable this odds ratio gives the ratio of the odds of the event in those with and without that feature. For continuous predictors it relates to the increase in odds associated with an increase of one unit in the value of the variable. Estimated log odds ratios from several studies can be combined by using the inverse variance method.²⁷

Outcome is time to event

When the time to event is explicitly considered for each individual in a study, the data are analysed with "survival analysis" methods—most often the log rank test for simple comparisons or Cox regression for analyses of multiple predictor variables or where one or more variables is continuous. By analogy with logistic regression discussed above, these analyses yield hazard ratios, which are similar to relative risks. Log rank statistics and log hazard ratios can be combined using the Peto method or the inverse variance method, respectively.²⁷

Practical difficulties are likely to make meta-analysis more difficult than the preceding explanation suggests. Most obviously the hazard ratio is not always explicitly presented for each study. Parmar et al described several methods of deriving estimates of the necessary statistics in a variety of situations.²⁸ For example, an estimate can be derived from the P value of the log rank test. They also explain how to estimate the standard errors of these estimates.

Several authors have proposed more complex methods for combining data from several studies of survival.^{29, 30} All can be applied in this context if it is possible to extract suitable data, but some require even more data than the basic items just discussed. The use of sophisticated statistical techniques may be inappropriate when several more basic weaknesses exist in the data. Indeed some reviewers have had to summarise the findings of the primary studies as P values as it is difficult to extract useful and usable quantitative information from many papers.³¹

Discussion

The principles of the systematic review should be extended to studies of prognosis, but doing so is by no means straightforward. The literature on prognosis features studies of poor quality and variable methodology, and the difficulties are exacerbated by

inadequate reporting of methodology. The poor quality of the published literature is a strong argument in favour of systematic reviews but also an argument against formal meta-analysis. To this end it is valuable if a systematic review includes details of the methodology of each study and its principal numerical results.³²

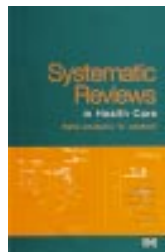
Although meta-analyses of published information may sometimes be useful, especially when the study characteristics do not vary too much and only the best studies are included, the findings are rarely convincing. The main outcome from such systematic reviews may be the realisation that there is little good quality information in the literature. Even an apparently clear result may best be seen as providing the justification for a well designed prospective study.³³

By contrast, meta-analysis based on individual patient data is highly desirable. Among several advantages of such data it is possible to analyse all the data in a consistent manner. It may also be possible to include data from unpublished studies. Meta-analysis of the raw data from all (or almost all) relevant studies is a worthy goal, and there have been some notable examples, especially in an epidemiological setting.³⁴ Apart from the considerable resources needed to carry out such a review, in most cases it is likely that many of the data sets are unobtainable. However, a careful collaborative reanalysis of the raw data from several good studies may be more valuable than a more superficial review that mixes good and poor studies. Two examples of such collaborative meta-analyses of raw data are a study of the relation between alcohol consumption and the development of breast cancer and a study of the relation between a vegetarian diet and mortality.^{35 36}

Poor quality studies may distort the results of a subsequent meta-analysis. When examined critically a high proportion of prognostic studies are found to be methodologically poor.^{9 37} Prognostic studies are generally too small and too poorly designed and analysed to provide reliable evidence. Although some suggested guidelines have appeared,^{2 13} progress may depend on developing a consensus regarding main methodological requirements for reliable studies of prognostic factors, as has happened for randomised trials.^{38 39}

As a consequence of the poor quality of research, prognostic markers may remain under investigation for many years after initial studies without any resolution of the uncertainty. Multiple separate and uncoordinated studies may actually delay the process of defining the role of prognostic markers. Systematic reviews can draw attention to the paucity of good quality evidence and, it is hoped, improve the quality of future research.

Competing interests: None declared.



Systematic Reviews in Health Care: Meta-analysis in Context can be purchased through the BMJ Bookshop (www.bmjbookshop.com); further information and updates for the book are available on www.systematicreviews.com

- Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat* 1998;52:289-303.
- Altman DG. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001:228-47.
- McKibbin K, Eady A, Marks S. *Evidence-based principles and practice*. PDQ Series. Hamilton, Canada: BC Dekker, 1999.
- Easterbrook PJ, Berlin J, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
- Shaheen NJ, Crosby MA, Bozynski ME, Sandler RS. Is there publication bias in the reporting of cancer risk in Barrett's esophagus. *Gastroenterology* 2000;119:333-8.

- Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;000:000-000.
- Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess* 1999;3(12):1-98.
- Shea B, Dubé C, Moher D. Assessing the quality of reports of systematic reviews: the QUOROM statement compared to other tools. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001:122-39.
- Kernan WN, Feinstein AR, Brass LM. A methodological appraisal of research on prognosis after transient ischemic attacks. *Stroke* 1991;22:1108-16.
- Levine MN, Browman GP, Gent M, Roberts R, Goodyear M. When is a prognostic factor useful?: a guide for the perplexed. *J Clin Oncol* 1991;9:348-56.
- Laupacis A, Wells G, Richardson S, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994;272:234-7.
- Hudak PL, Cole DC, Haines T. Understanding prognosis to improve rehabilitation: the example of lateral elbow pain. *Arch Phys Med Rehabil* 1996;77:586-93.
- Marx BE, Marx M. Prognosis of idiopathic membranous nephropathy: a methodologic meta-analysis. *Kidney Int* 1997;51:873-9.
- Low N, Egger M. Can we predict the prognosis of HIV infection? How to use the findings of a prospective study. *Sex Transm Infect* 1998;74:149-54.
- Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JHP, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
- Windeler J. Prognosis—what does the clinician associate with this notion? *Stat Med* 2000;19:425-30.
- Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;69:979-85.
- Brocklehurst P, French R. The association between maternal HIV infection and perinatal outcome: a systematic review of the literature and meta-analysis. *Br J Obstet Gynaecol* 1998;105:836-48.
- Wu YW, Colford JM. Chorioamnionitis as a risk factor for cerebral palsy. A meta-analysis. *JAMA* 2000;284:1417-24.
- Cohen J. The cost of dichotomization. *Appl Psychol Meas* 1983;7:249-53.
- Morgan TM, Elashoff RM. Effect of categorizing a continuous covariate on the comparison of survival time. *J Am Stat Assoc* 1986;81:917-21.
- Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994;86:829-35.
- Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* 1999;28:1-9.
- Stephenson JM. Systematic review of hormonal contraception and risk of HIV transmission: when to resist meta-analysis. *AIDS* 1998;12:545-53.
- Wang CC, Kreiss JK, Reilly M. Risk of HIV infection in oral contraceptive pill users: a meta-analysis. *J Acquir Immune Defic Syndr* 1999;21:51-8.
- Buettner P, Garbe C, Guggenmoos-Holzmann I. Problems in defining cutoff points of continuous prognostic factors: example of tumor thickness in primary cutaneous melanoma. *J Clin Epidemiol* 1997;50:1201-10.
- Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001:285-312.
- Parmar M, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998;17:2815-34.
- Hunink MGM, Wong JB. Meta-analysis of failure-time data with adjustment for covariates. *Med Decis Making* 1994;14:59-70.
- Dear KBG. Iterative generalised least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994;50:989-1002.
- Fox SB, Smith K, Hollyer J, Greenall M, Hastrich D, Harris AL. The epidermal growth factor receptor as a prognostic marker: results of 370 patients and review of 3009 patients. *Breast Cancer Res Treat* 1994;29:41-9.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF for the QUOROM Group. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999;354:1896-900.
- Ray JG. Meta-analysis of hyperhomocysteinemia as a risk factor for venous thromboembolic disease. *Arch Intern Med* 1998;158:2101-6.
- Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data of 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996;347:1713-27.
- Key TJ, Fraser GE, Thorogood M, Appleby PN, Beral V, Reeves G, et al. Mortality in vegetarians and nonvegetarians: detailed findings from a collaborative analysis of 5 prospective studies. *Am J Clin Nutr* 1999;70(suppl 3):S516-24.
- Smith-Warner SA, Spiegelman D, Yaun S-S, van den Brandt PA, Folsom AR, Goldbohm A, et al. Alcohol and breast cancer in women. A pooled analysis of cohort studies. *JAMA* 1998;279:535-40.
- Borghouts JAJ, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: a systematic review. *Pain* 1998;77:1-13.
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA* 1996;276:637-9.
- Moher D, Schulz KF, Altman DG for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.