

Evaluation of the Quality of Prognosis Studies in Systematic Reviews

Jill A. Hayden, DC; Pierre Côté, DC, PhD; and Claire Bombardier, MD

Background: To provide valid assessments of answers to prognostic questions, systematic reviews must appraise the quality of the available evidence. However, no standard quality assessment method is currently available.

Purpose: To appraise how authors assess the quality of individual studies in systematic reviews about prognosis and to propose recommendations for these quality assessments.

Data Sources: English-language publications listed in MEDLINE from 1966 to October 2005 and review of cited references.

Study Selection: 163 systematic reviews about prognosis that included assessments of the quality of studies.

Data Extraction: A total of 882 distinct quality items were extracted from the assessments that were reported in the various reviews. Using an iterative process, 2 independent reviewers grouped the items into 25 domains. The authors then specifically identified domains necessary to assess potential biases of studies and evaluated how often those domains had been addressed in each review.

Data Synthesis: Fourteen of the domains addressed 6 sources of bias related to study participation, study attrition, measurement of prognostic factors, measurement of and controlling for confounding variables, measurement of outcomes, and analysis approaches. Reviews assessed a median of 2 of the 6 potential biases; only 2 (1%) included criteria aimed at appraising all potential sources of bias. Few reviews adequately assessed the impact of confounding (12%), although more than half (59%) appraised the methods used to measure the prognostic factors of interest.

Limitations: Reviews may have been missed by the search or misclassified because of incomplete reporting. Validity and reliability testing of the authors' recommendations are required.

Conclusions: Quality appraisal, a necessary step in systematic reviews, is incomplete in most reviews of prognosis studies. Adequate quality assessment should include judgments about 6 areas of potential study biases. Authors should incorporate these quality assessments into their synthesis of evidence about prognosis.

Ann Intern Med. 2006;144:427-437.

www.annals.org

For author affiliations, see end of text.

Prognosis studies are investigations of future events or the evaluation of associations between risk factors and health outcomes in populations of patients (1). The results of such studies improve our understanding of the clinical course of a disease and assist clinicians in making informed decisions about how best to manage patients. Prognostic research also informs the design of intervention studies by helping define subgroups of patients who may benefit from a new treatment and by providing necessary information about the natural history of a disorder (2). There has recently been a rapid increase in the use of systematic review methods to synthesize the evidence on research questions related to prognosis.

It is essential that investigators conducting systematic reviews thoroughly appraise the methodologic quality of included studies to be confident that a study's design, conduct, analysis, and interpretation have adequately reduced the opportunity for bias (3, 4). Caution is warranted, however, because inclusion of methodologically weak studies can threaten the internal validity of a systematic review (4). This follows abundant empirical evidence that inadequate attention to biases can cause invalid results and inferences (5–9). However, there is limited consensus on how to appraise the quality of prognosis studies (1). A useful framework to assess bias in such studies follows the basic principles of epidemiologic research (10, 11). We focus on 6 areas of potential bias: study participation, study attrition, prognostic factor measurement, confounding measurement and account, outcome measurement, and analysis.

The main objectives of our "review of reviews" are to describe methods used to assess the quality of prognosis

studies and to describe how well current practices assess potential biases. Our secondary objective is to develop recommendations to guide future quality appraisal, both within single studies of prognostic factors and within systematic reviews of the evidence. We hope this work facilitates future discussion and research on biases in prognosis studies and systematic reviews.

METHODS

Literature Search and Study Selection

We identified systematic reviews of prognosis studies by searching MEDLINE (1966 to October 2005) using the search strategy recommended by McKibbin and colleagues (12). This strategy combines broad search terms for systematic reviews (*systematic review.mp; meta-analysis.mp*) and a sensitive search strategy for prognosis studies (*cohort, incidence, mortality, follow-up studies, prognos*, predict*, or course*). We also searched the reference lists of included reviews and methodologic papers to identify other relevant publications. We restricted our search to English-language publications. One reviewer conducted the search and selected the studies. Systematic reviews, defined as reviews of published studies with a comprehensive search and system-

See also:

Web-Only

Conversion of figures and tables into slides

atic selection, were included if they assessed the methodologic quality of the included studies by using 1 or more explicit criteria. We excluded studies if they were meta-analyses of independent patient data only, if their primary goal was to investigate the effectiveness of an intervention or specific diagnostic or screening tests, or if they included studies that were not done on humans.

Data Extraction and Synthesis

Individual items included in the quality assessment of the systematic reviews were recorded as they were reported in the publication (that is, the information that would be available to readers and future reviewers). We reviewed journal Web sites and contacted the authors of the systematic reviews for additional information when authors made such an offer in their original papers. When reviews assessed different study designs by using different sets of quality items, we extracted only those items used to assess cohort studies.

We constructed a comprehensive list of distinct items that the reviews used to assess the quality of their included studies. The full text of each review was screened. All items used by the review authors to assess the quality of studies were extracted into a computerized spreadsheet by 1 reviewer.

Two experienced reviewers, a clinical epidemiologist and an epidemiologist, independently synthesized the quality items extracted from the prognosis reviews to determine how well the systematic reviews assessed potential biases. We did this in 3 steps: 1) identified distinct concepts or domains addressed by the quality items; 2) grouped each extracted quality item into the appropriate domain or domains; and 3) identified the domains necessary to assess potential biases in prognosis studies. We then used this information to assess how well the reviews' quality assessment included items from the domains necessary to assess potential biases. After completing each of the first 3 steps, the reviewers met to attempt to reach a consensus. The consensus process involved each reviewer presenting his or her observations and results, followed by discussion and debate. A third reviewer was available in cases of persistent disagreement or uncertainty.

In the first step, all domains addressed by the quality items were identified. The first reviewer iteratively and progressively defined the domains as items were extracted from the included reviews. The second reviewer defined domains from a random list of all extracted quality items. Limited guidance was provided to the reviewers so that their assessments and definitions of domains would be independent. The reviewers agreed on a final set of domains that adequately and completely defined all of the extracted items.

In the second step, reviewers independently grouped each extracted item into the appropriate domains. Reviewers considered each extracted item by asking, "What is each particular quality item addressing?" or "What are the review's authors 'getting at' with the particular quality assess-

ment item?". Items were grouped into the domain or domains that best represented the concepts being addressed. For example, the extracted items "at least 80% of the group originally identified was located for follow-up" and "follow-up was sufficiently complete or doesn't jeopardize validity" were each independently classified by both reviewers as assessing the domain "completeness of follow-up adequate," whereas the extracted item "quantification and description of all subjects lost to follow-up" was classified as assessing the domain "completeness of follow-up described."

In the third step, we identified the domains necessary to assess potential biases. Each reviewer considered the ability of the identified domains to adequately address, at least in part, 1 of the following 6 potential biases: 1) study participation, 2) study attrition, 3) prognostic factor measurement, 4) confounding measurement and account, 5) outcome measurement, and 6) analysis. Domains were considered to adequately address part of the framework if information garnered from that domain would inform the assessment of potential bias. For example, both reviewers judged that the identified domain "study population represents source population or population of interest" assessed potential bias in a prognosis study, whereas the domain "research question definition" did not, although the latter is an important consideration in assessing the inclusion of studies in a systematic review.

Finally, on the basis of our previous ratings, we looked at whether each review included items from the domains necessary to assess the 6 potential biases. We calculated the frequency of systematic reviews by assessing each potential bias and the number of reviews that adequately assessed bias overall. From this systematic synthesis, we developed recommendations for improving quality appraisal in future systematic reviews of prognosis studies.

We used Microsoft Access and Excel 2002 (Microsoft Corp., Redmond, Washington) for data management and SAS for Windows, version 9.1 (SAS Institute, Inc., Cary, North Carolina) for descriptive statistics.

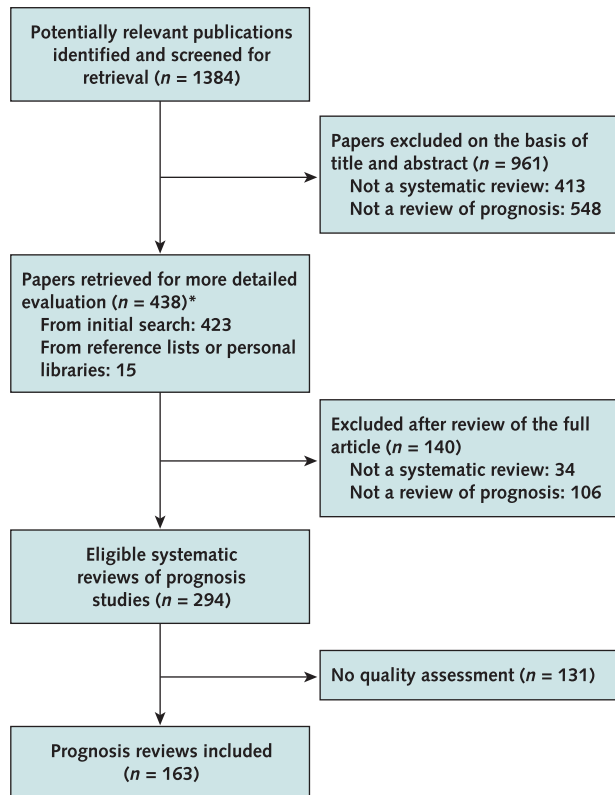
Role of the Funding Sources

The funding sources, the Canadian Institutes of Health Research, the Canadian Chiropractic Research Foundation, the Ontario Chiropractic Association, and the Ontario Ministry of Health and Long Term Care, did not have a role in the collection, analysis, or interpretation of the data or in the decision to submit the manuscript for publication.

RESULTS

We identified 1384 potentially relevant articles. **Figure 1** shows a flow chart of studies that were included and excluded. **Figure 2** shows the number of reviews identified by year of publication. We excluded 131 systematic reviews of prognosis studies that did not seem to include any quality assessment of the included studies; this represented 44% of prognosis reviews. We included 163 reviews of

Figure 1. Flow diagram of inclusion and exclusion criteria of systematic reviews.



The primary reason for exclusion is noted. *Includes 4 articles not available for full-article screening.

prognosis studies in our analysis (13–175). The most common topics were cancer (15%), musculoskeletal disorders and rheumatology (13%), cardiovascular (10%), neurology (10%), and obstetrics (10%). Other reviews included a wide range of health and health care topics. Sixty-three percent of the reviews investigated the association between a specific prognostic factor and a particular outcome; the remainder investigated multiple prognostic factors or models. The number of primary studies included in each systematic review ranged from 3 to 167 (median, 18 [interquartile range, 12 to 31]). A complete description of the included reviews is available from the authors on request.

Quality Items

One hundred fifty-three reviews provided adequate detail to allow extraction of quality items. Eight hundred eighty-two distinct quality items were extracted from the reviews. Most reviews developed their own set of quality items, with only a few applying criteria from previous reviews. Most quality items extracted from the reviews were not well defined. For example, “analyzed appropriately” and “controlled for confounding” were commonly used descriptors to assess the adequacy of the statistical analysis and of accounting for confounding, respectively. The qual-

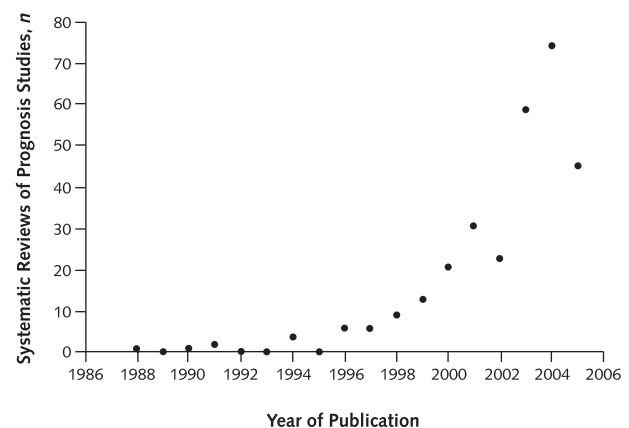
ity items included in the reviews also differed in the selected level or dose used. For example, for adequate follow-up, levels were often set and ranged between 65% (130, 139) and 95% (92, 105). Also, specific quality items assessed in the different reviews sometimes contradicted each other. For example, with respect to statistical analysis, the use of stepwise regression in the primary study was considered positive in 1 review (32), yet it was an exclusion criterion (that is, fatal flaw) in another review (31).

Assessment of Opportunities for Bias

The 2 reviewers independently identified 17 and 30 domains of quality items, respectively. The main difference between the reviewers was the explicitness of the definitions. For example, reviewer 1 identified “prognostic factor measured” and reviewer 2 identified “prognostic factor defined,” “prognostic factor adequate,” and “prognostic factor measured appropriately” to assess the same bias. After discussion, we reached consensus on 25 domains of quality items, which are shown in Table 1. The 2 reviewers initially agreed on grouping 652 of 882 quality items into the 25 domains (74% agreement). Most discrepancies were due to different interpretations of quality items by the 2 reviewers, which were resolved after discussion.

Fourteen of the 25 domains at least partly assessed 1 of the 6 potential biases. Adequate assessment of each bias required the review to judge how well each study’s methods limited the risk for the bias as opposed to the study simply reporting what methods were used (Table 2). Adequate assessment of biases was between 13% (for confounding measurement and account) and 59% (for prognostic factor measurement). Reviews adequately assessed a median of 2 (interquartile range, 1 to 4) of the 6 potential biases. Only 2 reviews adequately assessed all 6 biases (79, 80), and 21 reviews addressed 5 of the 6 (14, 23, 24, 26, 30–32, 34, 67, 72, 73, 93, 100–102, 116, 117, 141, 148,

Figure 2. Number of systematic reviews of prognosis studies identified over time.



The electronic search included 1996 to October 2005.

Table 1. Domains Abstracted from the Included Systematic Reviews

1. Source population clearly defined
2. Study population described
3. Study population represents source population or population of interest
4. Completeness of follow-up described
5. Completeness of follow-up adequate
6. Prognostic factors defined
7. Prognostic factors measured appropriately
8. Outcome defined
9. Outcome measured appropriately
10. Confounders defined and measured
11. Confounding accounted for
12. Analysis described
13. Analysis appropriate
14. Analysis provides sufficient presentation of data
15. External validation of results
16. Follow-up length appropriate
17. Follow-up length described
18. General appropriateness of outcome
19. General internal validity
20. Geographic area
21. Research question definition
22. Sample size adequate
23. Study design adequate
24. Year of publication
25. Evidence supporting conclusions

155, 171). Eleven domains from the reviews did not fit our framework because they were not judged to address potential biases (Table 1, items 15 to 25). These included domains relevant to the review research question, generalizability, evidence supporting the conclusions of the study, and domains used as general markers of overall quality. Fifty-four percent of reviews ($n = 82$) included domains in their quality assessment that were related to defining the research question. Seventy-one percent of reviews ($n = 109$) included at least 1 item that may be viewed as an overall marker of quality.

Integration of Study Quality in Systematic Reviews

Three main approaches were used to integrate study quality into the systematic reviews (some reviews used more than 1 approach). Fifty-five reviews used quality items as inclusion or exclusion criteria in the initial screening of eligible articles (18 reviews) or in a secondary screening phase as “fatal flaw” criteria (37 reviews). Forty reviews assessed overall study quality by assigning a quality score on the basis of a fixed number of items, followed by ranking (28 reviews) or use of a cutoff point (for example, < 50% of criteria satisfied) to exclude studies from the synthesis (12 reviews). Forty-two reviews tested the association of quality and outcome either by presenting results for subgroups of studies on the basis of methodologic quality (36 reviews: 16 based on overall quality score and 20 based on results of individual quality items) or by using meta-regression analysis (5 reviews: 3 based on overall quality score and 2 based on individual quality items). Finally, 24 reviews described the quality assessment; how-

ever, they did not incorporate study quality into the review synthesis.

Association between Methodologic Quality and Outcome

The 42 reviews that tested the association of quality and outcome potentially provide empirical evidence. The presentation of these data, however, was limited and the findings were inconsistent. Evans and Levene (53) found a statistically significant trend toward increased survival of infants born preterm in studies with higher risk for selection bias. Other reviews did not find an association between selection bias and outcome (37, 110, 111, 167, 170, 173, 174). Prognostic factor measurement bias was examined in 8 reviews with no reported clinically or statistically significant associations with results. Poor control of confounding was associated with inflated effect estimates in 4 studies: 1 on *Helicobacter pylori* infection and gastric cancer (40), 1 on laboratory variables and lung cancer (163), 1 on prevalence of specific conditions in carpal tunnel syndrome (157), and 1 on breastfeeding and childhood obesity (175). However, other reviews did not corroborate these findings (62, 64, 119, 127, 141, 167, 174). Four of 5 available studies did not find an association between adequate outcome measurements and the size of the prognostic effect (62, 64, 149, 173); O’Leary and colleagues (122) reported more stable results with longer-term outcome measurement in their review on primary affective disorders and suicide.

DISCUSSION

We conducted a review of systematic reviews of prognosis studies and observed a considerable increase in the number of such reviews published recently. We found, however, that quality assessment in the primary studies is often incomplete and that there is wide variation in current practice. Our results are similar to the findings of Deeks and colleagues (4), who looked at systematic reviews of intervention effectiveness, including nonrandomized studies. We think that our review clearly shows the need for guidelines on assessing the quality of prognosis studies. We have identified 4 distinct elements that are necessary to adequately assess the quality of such studies in systematic reviews: 1) operationalization of items to address potential opportunities for bias, 2) assessment of biases, 3) synthesizing the evidence, and 4) reporting results.

Operationalization of Items

We define operationalization of items as how individual quality items are described to inform reviewers’ assessment of potential bias. We found that few reviews fully operationalized their quality items. We understand that the vague criteria described in publications might have been used more explicitly by the review authors, and therefore we may have underestimated the completeness of their quality assessment. However, inadequate documentation of

Table 2. Domains Included in the Framework of Potential Biases and the Proportion of Reviews Assessing the Biases*

Potential Bias	Studies Adequately Assessing Bias, %†	Domains Addressed	Studies Assessing Domain, %
1. The study sample represents the population of interest on key characteristics, sufficient to limit potential bias to the results (study participation).	55	1. Source population clearly defined 2. Study population described 3. Study population represents source population or population of interest	50 21 50
2. Loss to follow-up (from sample to study population) is not associated with key characteristics, sufficient to limit potential bias (i.e., the study data adequately represent the sample) (study attrition).	42	4. Completeness of follow-up described 5. Completeness of follow-up adequate	19 42
3. The prognostic factor of interest is adequately measured in study participants to sufficiently limit potential bias (prognostic factor measurement).	59	6. Prognostic factors defined 7. Prognostic factors measured appropriately	31 59
4. The outcomes of interest are adequately measured in study participants to sufficiently limit potential bias (outcome measurement).	51	8. Outcome defined 9. Outcome measured appropriately	42 51
5. Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest (confounding measurement and account).	13	10. Confounders defined and measured 11. Confounding accounted for	21 53
6. The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid results (analysis).	33	12. Analysis described 13. Analysis appropriate 14. Analysis provides sufficient presentation of data	8 33 32

* Data are from 153 prognostic systematic reviews with quality items that could be extracted.

†Adequate assessment included 1) study participation: “source population clearly defined” and “study population described” or “study population represents source population or population of interest”; 2) study attrition: “completeness of follow-up adequate”; 3) prognostic factor measurement: “prognostic factors measured appropriately”; 4) outcome measurement: “outcome measured appropriately”; 5) confounding measurement and account: “confounders defined and measured” and “confounding accounted for”; and 6) analysis: “analysis appropriate.”

the operationalization limits the readers’ abilities to interpret methods and findings. We recommend that authors of systematic reviews make their operationalizations accessible in the published article, through journal Web sites, or through author contact.

To assist review authors in operationalizing their quality items, our recommendations include a comprehensive list of items that were extracted from the reviews of prognosis studies and supplemented by recent methodologic studies (Table 3). This list may include items that are irrelevant to some review questions; reviewers should specify a priori those items that are relevant to their question. For example, a review research question that considers the impact of multiple potential prognostic factors will need to consider potential confounders for each factor, whereas a review of 1 factor will need to consider only potential confounders for the 1 prognostic factor of interest. Our recommendations also require that review authors clarify some items on the basis of the review’s research question. For example, the item “all important confounders, including treatments, are measured” requires the reviewer to specify the important confounding variables. This decision should be based on previous research and a conceptual model.

Assessment of Biases

Assessment of biases refers to the assessment of potential biases within the studies included in a review. It was common for reviews to include quality items that assessed

the “reporting” of a study method rather than assessing how well the study’s methods limited bias. For example, some reviews included quality items assessing “outcome defined” rather than “outcome measured appropriately.” Good reporting in publications is important to assess study quality; however, this alone is not sufficient to assess risk for bias. In cases where individual biases were adequately assessed, few reviews comprehensively assessed all important biases. We recommend that quality appraisal of prognosis studies consider each of 6 potential biases (Table 3) in 2 steps. Step 1 is to assess the fully operationalized, relevant quality items (Table 3, column 2). These item responses are then used to inform step 2, where each of the 6 potential biases (Table 3, column 1) are judged. This approach to quality appraisal follows a method that has been described by Wortman (176) as “mixed-criteria” quality assessment. We recommend against “quality score” approaches that assign points on the basis of the number of “positive” quality items because this reduces scientific judgment. Review authors are also discouraged from setting explicit values to assess risk for bias. For example, an item “loss to follow-up less than 20%” is not adequate to assess risk for bias associated with study attrition. Instead, we suggest reviewers thoughtfully consider overlapping methodologic issues and the direction of the potential bias for each case. For example, low risk for attrition bias would require adequate follow-up as well as demonstration that the response is not likely to be associated with the prog-

Table 3. Guidelines for Assessing Quality in Prognostic Studies on the Basis of Framework of Potential Biases*

Potential Bias	Items To Be Considered for Assessment of Potential Opportunity for Bias
Study participation The study sample represents the population of interest on key characteristics, sufficient to limit potential bias to the results. Yes Partly No Unsure	The source population or population of interest is adequately described for key characteristics. The sampling frame and recruitment are adequately described, possibly including methods to identify the sample (number and type used, e.g., referral patterns in health care), period of recruitment, and place of recruitment (setting and geographic location) Inclusion and exclusion criteria are adequately described (e.g., including explicit diagnostic criteria or "zero time" description). There is adequate participation in the study by eligible individuals. The baseline study sample (i.e., individuals entering the study) is adequately described for key characteristics.
Study attrition Loss to follow-up (from sample to study population) is not associated with key characteristics (i.e., the study data adequately represent the sample), sufficient to limit potential bias. Yes Partly No Unsure	Response rate (i.e., proportion of study sample completing the study and providing outcome data) is adequate. Attempts to collect information on participants who dropped out of the study are described. Reasons for loss to follow-up are provided. Participants lost to follow-up are adequately described for key characteristics. There are no important differences between key characteristics and outcomes in participants who completed the study and those who did not.
Prognostic factor measurement The prognostic factor of interest is adequately measured in study participants to sufficiently limit potential bias. Yes Partly No Unsure	A clear definition or description of the prognostic factor measured is provided (e.g., including dose, level, duration of exposure, and clear specification of the method of measurement). Continuous variables are reported or appropriate (i.e., not data-dependent) cut-points are used. The prognostic factor measure and method are adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall). Adequate proportion of the study sample has complete data for prognostic factors. The method and setting of measurement are the same for all study participants. Appropriate methods are used if imputation is used for missing prognostic factor data.
Outcome measurement The outcome of interest is adequately measured in study participants to sufficiently limit potential bias. Yes Partly No Unsure	A clear definition of the outcome of interest is provided, including duration of follow-up and level and extent of the outcome construct. The outcome measure and method used are adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test). The method and setting of measurement are the same for all study participants.
Confounding measurement and account Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest. Yes Partly No Unsure	All important confounders, including treatments (key variables in conceptual model), are measured. Clear definitions of the important confounders measured are provided (e.g., including dose, level, and duration of exposures). Measurement of all important confounders is adequately valid and reliable (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall). The method and setting of confounding measurement are the same for all study participants. Appropriate methods are used if imputation is used for missing confounder data. Important potential confounders are accounted for in the study design (e.g., matching for key variables, stratification, or initial assembly of comparable groups). Important potential confounders are accounted for in the analysis (i.e., appropriate adjustment).
Analysis The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid results. Yes Partly No Unsure	There is sufficient presentation of data to assess the adequacy of the analysis. The strategy for model building (i.e., inclusion of variables) is appropriate and is based on a conceptual framework or model. The selected model is adequate for the design of the study. There is no selective reporting of results.

* Guidelines should be applied on the basis of relevance to the review research question.

nostic factor and outcome of interest (that is, random loss to follow-up). This thoughtful approach to assessing risk for bias will require trained data extractors or readers. Furthermore, the assessment of risk for bias should be

completed by at least 2 independent reviewers. Reviewers' consensus process should include discussion and debate of the independent ratings of each item and the overall risk for each potential bias.

Synthesizing the Evidence

Synthesizing the evidence describes how the results of quality assessment are incorporated into a review's synthesis of the evidence. Although some reviews assessed quality, not all used the results of their assessments when interpreting the evidence. We recommend including information on each of the 6 biases separately in the review synthesis (that is, component-based analysis). For example, the evidence of effect should be presented on the basis of studies with low risk for bias associated with study participation, study attrition, prognostic factor measurement, outcome measurement, confounding measurement and account, and analysis.

It may also be desirable for a review synthesis to include an assessment of evidence of effect based on studies with an overall low risk for any important bias. Therefore, we suggest that studies of acceptable quality for inclusion in the synthesis would at least partly satisfy each of the 6 biases (that is, studies from the analysis that are at high risk for any important bias would be omitted). Sensitivity analyses should be used in review syntheses to explore the impact of choices regarding study quality.

Reporting Results

Finally, we highlight the need to adequately report the results of quality assessment for each individual study and also for the group of included studies. This ensures the transparency of the systematic review and allows readers to interpret the results. Furthermore, this will provide empirical evidence of association of potential biases with outcome. Reporting such results will help users of research—clinicians, policymakers, and funders—better understand the impact of bias on prognosis studies and may yield better-quality research over the long term.

Limitations and Next Steps

We think our review of systematic reviews of prognosis studies has accomplished our objectives of describing current quality assessment practice and making recommendations for future systematic reviews. However, our work has several limitations and we encourage future discussion and research on assessing bias in prognosis studies.

In our study, we used systematic review methods and collected data from published reviews. Our methods to identify publications may have missed some eligible publications. Our search strategy yielded reviews published in English-language, peer-reviewed, MEDLINE-indexed journals and those with adequate assignment of keywords and Medical Subject Heading (MeSH) terms. The included reviews may be of higher quality than unidentified reviews published in non-English-language journals and in journals without peer review. However, this probably did not have an impact on our recommendations: The 2 reviewers reached saturation on domain generation (that is, no new domains were being generated from systematic review quality criteria). Therefore, we believe our sample was adequate and allowed us to make conclusions.

Second, we collected data on whether and how the reviewers assessed the quality of prognosis studies mainly from information available in the publications. We were therefore limited by incomplete reporting of the systematic review methods. Poor reporting of quality assessment that was actually done by the reviews may have led us to underestimate the extent and comprehensiveness of quality assessment in prognosis reviews.

Third, we had only 2 reviewers reach consensus on the identification of domains and the classification of quality items. It would have been ideal to conduct a formal consensus process with many international researchers. However, the results of the reviewers' independent classifications were similar at each step of our investigation. We do not think that a more comprehensive approach would substantially change our observation that quality assessment in systematic reviews of prognosis is often incomplete.

Finally, our recommendations for quality assessment require formal testing. The approach we used to develop our recommendations—starting with basic epidemiologic principles of bias and critically synthesizing the methods of many systematic reviews—provides minimal face and content validity. However, additional research is needed to apply our recommendations to systematic reviews in various topics and to test all aspects of validity and reliability. Testing of our recommendations is ongoing in 3 large prognosis reviews.

Conclusions

Systematic reviews of the evidence are useful tools for those who wish to identify, evaluate, and summarize information on health care topics. Quality appraisal of relevant studies is an important part of systematic reviews. However, we found that in reviews of prognosis, quality was often incompletely assessed. For prognosis reviews, we recommend complete and clear operationalization of quality items focused on the assessment of potential biases. Results of quality assessment should be included in the summary of the evidence and should be reported to allow interpretation by readers. We hope our work provides some groundwork and facilitates future discussion and research on the impact of biases in prognosis studies and systematic reviews.

From Institute for Work and Health, University of Toronto, and Mount Sinai Hospital, Toronto, Ontario, Canada.

Acknowledgments: The authors thank Drs. Doug Altman and Greta Ridley for valuable comments on earlier versions of this manuscript, Drs. Sheilah Hogg-Johnson and George Tomlinson for comments and discussion on the design and analysis, Evelyne Michaels for her assistance with editing and useful comments on an earlier draft of this manuscript, and Jeremy Dacombe and Quenby Mahood for assistance with literature retrieval.

Grant Support: This project was partially funded by a research grant from the Ontario Chiropractic Association and Ontario Ministry of Health and Long Term Care Special Chiropractic Research Fund. Dr.

Hayden is supported by a Postdoctoral Fellowship Award from the Canadian Institutes of Health Research and Canadian Chiropractic Research Foundation. Dr. Côté is supported by a New Investigator Award from the Canadian Institutes for Health Research and by the Institute for Work and Health and the Workplace Safety and Insurance Board of Ontario. Dr. Bombardier holds a Canada Research Chair in Knowledge Transfer for Musculoskeletal Care.

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Jill A. Hayden, DC, Institute for Work and Health, 481 University Avenue, Suite 800, 8th Floor, Toronto, Ontario M5G 2E9, Canada; e-mail, jhayden@iwh.on.ca.

Current author addresses are available at www.annals.org.

References

- Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ*. 2001;323:224-8. [PMID: 11473921]
- Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat*. 1998;52:289-303. [PMID: 10066088]
- Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490. [PMID: 15205295]
- Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7:iii-x, 1-173. [PMID: 14499048]
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615-25. [PMID: 15308962]
- Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*. 1991;134:1233-44. [PMID: 1746532]
- Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health*. 2001;22:189-212. [PMID: 11274518]
- Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med*. 2004;66:868-75. [PMID: 15564351]
- Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol*. 2004;19:751-60. [PMID: 15469032]
- Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51-63. [PMID: 447779]
- Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58:635-41. [PMID: 15252064]
- McKibbin K, Eady A, Marks S. Evidence Based Principles and Practice. PDQ Series. Hamilton, Canada: BC Dekker; 1999.
- Bonell C, Weatherburn P, Hickson F. Sexually transmitted infection as a risk factor for homosexual HIV transmission: a systematic review of epidemiological studies. *Int J STD AIDS*. 2000;11:697-700. [PMID: 11089782]
- Bongers PM, Kremer AM, ter Laak J. Are psychosocial factors, risk factors for symptoms and signs of the shoulder, elbow, or hand/wrist?: a review of the epidemiological literature. *Am J Ind Med*. 2002;41:315-42. [PMID: 12071487]
- Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: a systematic review. *Pain*. 1998;77:1-13. [PMID: 9755013]
- 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. [PMID: 9746375]
- Brooks RG, Walsh M, Mardon RE, Lewis M, Clawson A. The roles of nature and nurture in the recruitment and retention of primary care physicians in rural areas: a review of the literature. *Acad Med*. 2002;77:790-8. [PMID: 12176692]
- Burt BA, Pai S. Does low birthweight increase the risk of caries? A systematic review. *J Dent Educ*. 2001;65:1024-7. [PMID: 11699973]
- Campbell SE, Seymour DG, Primrose WR. A systematic literature review of factors affecting outcome in older medical patients admitted to hospital. *Age Ageing*. 2004;33:110-5. [PMID: 14960424]
- Cantor-Graae E, Selten JP. Schizophrenia and migration: a meta-analysis and review. *Am J Psychiatry*. 2005;162:12-24. [PMID: 15625195]
- Capes SE, Hunt D, Malmberg K, Gerstein HC. Stress hyperglycaemia and increased risk of death after myocardial infarction in patients with and without diabetes: a systematic overview. *Lancet*. 2000;355:773-8. [PMID: 10711923]
- Capes SE, Hunt D, Malmberg K, Pathak P, Gerstein HC. Stress hyperglycaemia and prognosis of stroke in nondiabetic and diabetic patients: a systematic overview. *Stroke*. 2001;32:2426-32. [PMID: 11588337]
- Carroll LJ, Cassidy JD, Peloso PM, Borg J, von Holst H, Holm L, et al. Prognosis for mild traumatic brain injury: results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *J Rehabil Med*. 2004;84-105. [PMID: 15083873]
- Carter JA, Neville BG, Newton CR. Neuro-cognitive impairment following acquired central nervous system infections in childhood: a systematic review. *Brain Res Brain Res Rev*. 2003;43:57-69. [PMID: 14499462]
- Chow E, Harth T, Hruby G, Finkelstein J, Wu J, Danjoux C. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. *Clin Oncol (R Coll Radiol)*. 2001;13:209-18. [PMID: 11527298]
- Cole DC, Hudak PL. Prognosis of nonspecific work-related musculoskeletal disorders of the neck and upper extremity. *Am J Ind Med*. 1996;29:657-68. [PMID: 8773726]
- Cole MG, Dendukuri N. Risk factors for depression among elderly community subjects: a systematic review and meta-analysis. *Am J Psychiatry*. 2003;160:1147-56. [PMID: 12777274]
- Cole MG, Bellavance F. The prognosis of depression in old age. *Am J Geriatr Psychiatry*. 1997;5:4-14. [PMID: 9169240]
- Conde-Agudelo A, Althabe F, Belizán JM, Kafury-Goeta AC. Cigarette smoking during pregnancy and risk of preeclampsia: a systematic review. *Am J Obstet Gynecol*. 1999;181:1026-35. [PMID: 10521771]
- Coo H, Aronson KJ. A systematic review of several potential non-genetic risk factors for multiple sclerosis. *Neuroepidemiology*. 2004;23:1-12. [PMID: 14739563]
- Côté P, Cassidy JD, Carroll L, Frank JW, Bombardier C. A systematic review of the prognosis of acute whiplash and a new conceptual framework to synthesize the literature. *Spine*. 2001;26:E445-58. [PMID: 11698904]
- Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis*. 2001;12:159-70. [PMID: 11641579]
- Counsell CE, Grant R. Incidence studies of primary and secondary intracranial tumors: a systematic review of their methodology and results. *J Neurooncol*. 1998;37:241-50. [PMID: 9524082]
- Coventry PA, Grande GE, Richards DA, Todd CJ. Prediction of appropriate timing of palliative care for older adults with non-malignant life-threatening disease: a systematic review. *Age Ageing*. 2005;34:218-27. [PMID: 15863407]
- Critchley JA, Unal B. Health effects associated with smokeless tobacco: a systematic review. *Thorax*. 2003;58:435-43. [PMID: 12728167]
- Critchley JA, Unal B. Is smokeless tobacco a risk factor for coronary heart disease? A systematic review of epidemiological studies. *Eur J Cardiovasc Prev Rehabil*. 2004;11:101-12. [PMID: 15187813]
- Critchley JA, Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. *JAMA*. 2003;290:86-97. [PMID: 12837716]
- Curtis V, Cairncross S. Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *Lancet Infect Dis*. 2003;3:275-81. [PMID: 12726975]
- Damany DS, Parker MJ, Chojnowski A. Complications after intracapsular hip fractures in young adults. A meta-analysis of 18 published studies involving 564 fractures. *Injury*. 2005;36:131-41. [PMID: 15589931]
- Danesh J. *Helicobacter pylori* infection and gastric cancer: systematic review of the epidemiological studies. *Aliment Pharmacol Ther*. 1999;13:851-6. [PMID: 10383517]
- Davies G, Welham J, Chant D, Torrey EF, McGrath J. A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr Bull*. 2003;29:587-93. [PMID: 14609251]
- de Bock GH, Bonnema J, van der Hage J, Kievit J, van de Velde CJ. Effectiveness of routine visits and routine tests in detecting isolated locoregional recurrences after treatment for early-stage invasive breast cancer: a meta-analysis and systematic review. *J Clin Oncol*. 2004;22:4010-8. [PMID: 15459225]
- Devereaux PJ, Schönemann HJ, Ravindran N, Bhandari M, Garg AX, Choi PT, et al. Comparison of mortality between private for-profit and private

- not-for-profit hemodialysis centers: a systematic review and meta-analysis. *JAMA*. 2002;288:2449-57. [PMID: 12435258]
44. Ding X, Yang LY, Huang GW, Yang JQ, Liu HL, Wang W, et al. Role of AFP mRNA expression in peripheral blood as a predictor for postsurgical recurrence of hepatocellular carcinoma: a systematic review and meta-analysis. *World J Gastroenterol*. 2005;11:2656-61. [PMID: 15849829]
45. Douketis JD, Kearon C, Bates S, Duku EK, Ginsberg JS. Risk of fatal pulmonary embolism in patients with treated venous thromboembolism. *JAMA*. 1998;279:458-62. [PMID: 9466640]
46. Doust JA, Pietrzak E, Dobson A, Glasziou P. How well does B-type natriuretic peptide predict death and cardiac events in patients with heart failure: systematic review. *BMJ*. 2005;330:625. [PMID: 15774989]
47. Duckitt K, Harrington D. Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. *BMJ*. 2005;330:565. [PMID: 15743856]
48. Dwyer T, Couper D, Walter SD. Sources of heterogeneity in the meta-analysis of observational studies: the example of SIDS and sleeping position. *J Clin Epidemiol*. 2001;54:440-7. [PMID: 11337206]
49. El-Wakeel H, Umpleby HC. Systematic review of fibroadenoma as a risk factor for breast cancer. *Breast*. 2003;12:302-7. [PMID: 14659144]
50. Emery CA. Risk factors for injury in child and adolescent sport: a systematic review of the literature. *Clin J Sport Med*. 2003;13:256-68. [PMID: 12855930]
51. Espallargues M, Sampietro-Colom L, Estrada MD, Solà M, del Rio L, Setoain J, et al. Identifying bone-mass-related risk factors for fracture to guide bone densitometry measurements: a systematic review of the literature. *Osteoporos Int*. 2001;12:811-22. [PMID: 11716183]
52. Etminan M, Takkouche B, Isorna FC, Samii A. Risk of ischaemic stroke in people with migraine: systematic review and meta-analysis of observational studies. *BMJ*. 2005;330:63. [PMID: 15596418]
53. Evans DJ, Levene MI. Evidence of selection bias in preterm survival studies: a systematic review. *Arch Dis Child Fetal Neonatal Ed*. 2001;84:F79-84. [PMID: 11207220]
54. Farrell T, Chien PF, Gordon A. Intrapartum umbilical artery Doppler velocimetry as a predictor of adverse perinatal outcome: a systematic review. *Br J Obstet Gynaecol*. 1999;106:783-92. [PMID: 10453827]
55. Flossmann E, Schulz UG, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke*. 2004;35:212-27. [PMID: 14684773]
56. Flynn CA, Helwig AL, Meurer LN. Bacterial vaginosis in pregnancy and the risk of prematurity: a meta-analysis. *J Fam Pract*. 1999;48:885-92. [PMID: 10907626]
57. Ford ES, Smith SJ, Stroup DF, Steinberg KK, Mueller PW, Thacker SB. Homocyst(e)ine and cardiovascular disease: a systematic review of the evidence with special emphasis on case-control studies and nested case-control studies. *Int J Epidemiol*. 2002;31:59-70. [PMID: 11914295]
58. Fowkes FJ, Price JF, Fowkes FG. Incidence of diagnosed deep vein thrombosis in the general population: systematic review. *Eur J Vasc Endovasc Surg*. 2003;25:1-5. [PMID: 12525804]
59. Fowler-Brown A, Pignone M, Pletcher M, Tice JA, Sutton SF, Lohr KN, et al. Exercise tolerance testing to screen for coronary heart disease: a systematic review for the technical support for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2004;140:W9-24. [PMID: 15069009]
60. Fraser AB, Grimes DA. Effect of lactation on maternal body weight: a systematic review. *Obstet Gynecol Surv*. 2003;58:265-9. [PMID: 12665706]
61. Garg AX, Suri RS, Barrowman N, Rehman F, Matsell D, Rosas-Arellano MP, et al. Long-term renal prognosis of diarrhea-associated hemolytic uremic syndrome: a systematic review, meta-analysis, and meta-regression. *JAMA*. 2003;290:1360-70. [PMID: 12966129]
62. Gdalevich M, Mimouni D, David M, Mimouni M. Breast-feeding and the onset of atopic dermatitis in childhood: a systematic review and meta-analysis of prospective studies. *J Am Acad Dermatol*. 2001;45:520-7. [PMID: 11568741]
63. Ghali WA, Wasil BI, Brant R, Exner DV, Cornuz J. Atrial flutter and the risk of thromboembolism: a systematic review and meta-analysis. *Am J Med*. 2005;118:101-7. [PMID: 15694889]
64. González MA, Rodríguez Artalejo F, Calero JR. Relationship between socioeconomic status and ischaemic heart disease in cohort and case-control studies: 1960-1993. *Int J Epidemiol*. 1998;27:350-8. [PMID: 9698119]
65. Guise JM, McDonagh MS, Osterweil P, Nygren P, Chan BK, Helfand M. Systematic review of the incidence and consequences of uterine rupture in women with previous caesarean section. *BMJ*. 2004;329:19-25. [PMID: 15231616]
66. Harris R, Nicoll AD, Adair PM, Pine CM. Risk factors for dental caries in young children: a systematic review of the literature. *Community Dent Health*. 2004;21:71-85. [PMID: 15072476]
67. Harvie M, Hooper L, Howell AH. Central obesity and breast cancer risk: a systematic review. *Obes Rev*. 2003;4:157-73. [PMID: 12916817]
68. Haw C, Hawton K, Sutton L, Sinclair J, Deeks J. Schizophrenia and deliberate self-harm: a systematic review of risk factors. *Suicide Life Threat Behav*. 2005;35:50-62. [PMID: 15843323]
69. Hemels ME, Einarson A, Koren G, Lanctôt KL, Einarson TR. Antidepressant use during pregnancy and the rates of spontaneous abortions: a meta-analysis. *Ann Pharmacother*. 2005;39:803-9. [PMID: 15784808]
70. Hemingway H, Marmot M. Evidence based cardiology: psychosocial factors in the aetiology and prognosis of coronary heart disease. Systematic review of prospective cohort studies. *BMJ*. 1999;318:1460-7. [PMID: 10346775]
71. Hendricks HT, van Limbeek J, Geurts AC, Zwarts MJ. Motor recovery after stroke: a systematic review of the literature. *Arch Phys Med Rehabil*. 2002;83:1629-37. [PMID: 12422337]
72. Hendricks HT, Zwarts MJ, Plat EF, van Limbeek J. Systematic review for the early prediction of motor and functional outcome after stroke by using motor-evoked potentials. *Arch Phys Med Rehabil*. 2002;83:1303-8. [PMID: 12235613]
73. Hermens ML, van Hout HP, Terluin B, van der Windt DA, Beekman AT, van Dyck R, et al. The prognosis of minor depression in the general population: a systematic review. *Gen Hosp Psychiatry*. 2004;26:453-62. [PMID: 15567211]
74. Aveyard P, Adab P, Cheng KK, Wallace DM, Hey K, Murphy MF. Does smoking status influence the prognosis of bladder cancer? A systematic review. *BJU Int*. 2002;90:228-39. [PMID: 12133057]
75. Herpertz S, Kielmann R, Wolf AM, Hebebrand J, Senf W. Do psychosocial variables predict weight loss or mental health after obesity surgery? A systematic review. *Obes Res*. 2004;12:1554-69. [PMID: 15536219]
76. Hodgkinson B, Evans D, Wood J. Maintaining oral hydration in older adults: a systematic review. *Int J Nurs Pract*. 2003;9:S19-28. [PMID: 12801253]
77. Honest H, Bachmann LM, Sengupta R, Gupta JK, Kleijnen J, Khan KS. Accuracy of absence of fetal breathing movements in predicting preterm birth: a systematic review. *Ultrasound Obstet Gynecol*. 2004;24:94-100. [PMID: 15229924]
78. Honest H, Bachmann LM, Ngai C, Gupta JK, Kleijnen J, Khan KS. The accuracy of maternal anthropometry measurements as predictor for spontaneous preterm birth—a systematic review. *Eur J Obstet Gynecol Reprod Biol*. 2005;119:11-20. [PMID: 15734079]
79. Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM. Physical load during work and leisure time as risk factors for back pain. *Scand J Work Environ Health*. 1999;25:387-403. [PMID: 10569458]
80. Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM. Systematic review of psychosocial factors at work and private life as risk factors for back pain. *Spine*. 2000;25:2114-25. [PMID: 10954644]
81. Hooi JD, Stoffers HE, Knottnerus JA, van Ree JW. The prognosis of non-critical limb ischaemia: a systematic review of population-based evidence. *Br J Gen Pract*. 1999;49:49-55. [PMID: 10622019]
82. Howard AA, Arnsten JH, Gourevitch MN. Effect of alcohol consumption on diabetes mellitus: a systematic review. *Ann Intern Med*. 2004;140:211-9. [PMID: 14757619]
83. Howley HE, Walker M, Rodger MA. A systematic review of the association between factor V Leiden or prothrombin gene variant and intrauterine growth restriction. *Am J Obstet Gynecol*. 2005;192:694-708. [PMID: 15746660]
84. Hughes EG, Brennan BG. Does cigarette smoking impair natural or assisted fecundity? *Fertil Steril*. 1996;66:679-89. [PMID: 8893667]
85. Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med*. 2002;137:273-84. [PMID: 12186518]
86. Jefferson T, Rudin M, DiPietrantonj C. Systematic review of the effects of pertussis vaccines in children. *Vaccine*. 2003;21:2003-14. [PMID: 12706690]
87. Kanaya AM, Grady D, Barrett-Connor E. Explaining the sex difference in coronary heart disease mortality among patients with type 2 diabetes mellitus: a meta-analysis. *Arch Intern Med*. 2002;162:1737-45. [PMID: 12153377]
88. Kernan WN, Feinstein AR, Brass LM. A methodological appraisal of research on prognosis after transient ischemic attacks. *Stroke*. 1991;22:1108-16. [PMID: 1926253]
89. Khader YS, Ta'ani Q. Periodontal diseases and the risk of preterm birth and

- low birth weight: a meta-analysis. *J Periodontol*. 2005;76:161-5. [PMID: 15974837]
90. Khanna MP, Hébert PC, Fergusson DA. Review of the clinical practice literature on patient characteristics associated with perioperative allogeneic red blood cell transfusion. *Transfus Med Rev*. 2003;17:110-9. [PMID: 12733104]
91. Kim RJ, Becker RC. Association between factor V Leiden, prothrombin G20210A, and methylenetetrahydrofolate reductase C677T mutations and events of the arterial circulatory system: a meta-analysis of published studies. *Am Heart J*. 2003;146:948-57. [PMID: 14660985]
92. Kremer LC, van Dalen EC, Offringa M, Voûte PA. Frequency and risk factors of anthracycline-induced clinical heart failure in children: a systematic review. *Ann Oncol*. 2002;13:503-12. [PMID: 12056699]
93. Kuijpers T, van der Windt DA, van der Heijden GJ, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. *Pain*. 2004;109:420-31. [PMID: 15157703]
94. Kuo HK, Yen CJ, Chang CH, Kuo CK, Chen JH, Sorond F. Relation of C-reactive protein to stroke, cognitive disorders, and depression in the general population: systematic review and meta-analysis. *Lancet Neurol*. 2005;4:371-80. [PMID: 15907742]
95. Kurihara N, Wada O. Silicosis and smoking strongly increase lung cancer risk in silica-exposed workers. *Ind Health*. 2004;42:303-14. [PMID: 15295901]
96. Kwon HL, Belanger K, Bracken MB. Effect of pregnancy and stage of pregnancy on asthma severity: a systematic review. *Am J Obstet Gynecol*. 2004;190:1201-10. [PMID: 15167819]
97. Langlois NJ, Wells PS. Risk of venous thromboembolism in relatives of symptomatic probands with thrombophilia: a systematic review. *Thromb Haemost*. 2003;90:17-26. [PMID: 12876621]
98. Lanigan JA, Bishop J, Kimber AC, Morgan J. Systematic review concerning the age of introduction of complementary foods to the healthy full-term infant. *Eur J Clin Nutr*. 2001;55:309-20. [PMID: 11378803]
99. LeBlanc ES, Janowsky J, Chan BK, Nelson HD. Hormone replacement therapy and cognition: systematic review and meta-analysis. *JAMA*. 2001;285:1489-99. [PMID: 11255426]
100. Lieveense A, Bierma-Zeinstra S, Verhagen A, Verhaar J, Koes B. Influence of work on the development of osteoarthritis of the hip: a systematic review. *J Rheumatol*. 2001;28:2520-8. [PMID: 11708427]
101. Lieveense AM, Bierma-Zeinstra SM, Verhagen AP, Verhaar JA, Koes BW. Influence of hip dysplasia on the development of osteoarthritis of the hip. *Ann Rheum Dis*. 2004;63:621-6. [PMID: 15140766]
102. Lieveense AM, Bierma-Zeinstra SM, Verhagen AP, Verhaar JA, Koes BW. Prognostic factors of progress of hip osteoarthritis: a systematic review. *Arthritis Rheum*. 2002;47:556-62. [PMID: 12382307]
103. Lin SJ, Schranz J, Teutsch SM. Aspergillosis case-fatality rate: systematic review of the literature. *Clin Infect Dis*. 2001;32:358-66. [PMID: 11170942]
104. Maltoni M, Caraceni A, Brunelli C, Broeckaert B, Christakis N, Eychmueller S, et al. Prognostic factors in advanced cancer patients: evidence-based clinical recommendations—a study by the Steering Committee of the European Association for Palliative Care. *J Clin Oncol*. 2005;23:6240-8. [PMID: 16135490]
105. Marckmann P, Grønbaek M. Fish consumption and coronary heart disease mortality. A systematic review of prospective cohort studies. *Eur J Clin Nutr*. 1999;53:585-90. [PMID: 10477243]
106. Marras C, Rochon P, Lang AE. Predicting motor decline and disability in Parkinson disease: a systematic review. *Arch Neurol*. 2002;59:1724-8. [PMID: 12433259]
107. Martin B, Paesmans M, Berghmans T, Branle F, Ghisdal L, Mascaux C, et al. Role of Bcl-2 as a prognostic factor for survival in lung cancer: a systematic review of the literature with meta-analysis. *Br J Cancer*. 2003;89:55-64. [PMID: 12838300]
108. Marx BE, Marx M. Prognosis of idiopathic membranous nephropathy: a methodologic meta-analysis. *Kidney Int*. 1997;51:873-9. [PMID: 9067924]
109. Mascaux C, Iannino N, Martin B, Paesmans M, Berghmans T, Dusart M, et al. The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. *Br J Cancer*. 2005;92:131-9. [PMID: 15597105]
110. McDonald S, Murphy K, Beyene J, Ohlsson A. Perinatal outcomes of in vitro fertilization twins: a systematic review and meta-analyses. *Am J Obstet Gynecol*. 2005;193:141-52. [PMID: 16021072]
111. McDonald SD, Murphy K, Beyene J, Ohlsson A. Perinatal outcomes of singleton pregnancies achieved by in vitro fertilization: a systematic review and meta-analysis. *J Obstet Gynaecol Can*. 2005;27:449-59. [PMID: 16100639]
112. Meert AP, Martin B, Paesmans M, Berghmans T, Mascaux C, Verdebout JM, et al. The role of HER-2/neu expression on the survival of patients with lung cancer: a systematic review of the literature. *Br J Cancer*. 2003;89:959-65. [PMID: 12966408]
113. Meijer R, Ihnenfeldt DS, de Groot IJ, van Limbeek J, Vermeulen M, de Haan RJ. Prognostic factors for ambulation and activities of daily living in the subacute phase after stroke. A systematic review of the literature. *Clin Rehabil*. 2003;17:119-29. [PMID: 12625651]
114. Meijer R, Ihnenfeldt DS, van Limbeek J, Vermeulen M, de Haan RJ. Prognostic factors in the subacute phase after stroke for the future residence after six months to one year. A systematic review of the literature. *Clin Rehabil*. 2003;17:512-20. [PMID: 12952157]
115. Meijer R, van Limbeek J, Kriek B, Ihnenfeldt D, Vermeulen M, de Haan R. Prognostic social factors in the subacute phase after a stroke for the discharge destination from the hospital stroke-unit. A systematic review of the literature. *Disabil Rehabil*. 2004;26:191-7. [PMID: 15164952]
116. Mondloch MV, Cole DC, Frank JW. Does how you do depend on how you think you'll do? A systematic review of the evidence for a relation between patients' recovery expectations and health outcomes. *CMAJ*. 2001;165:174-9. [PMID: 11501456]
117. Monteiro PO, Victora CG. Rapid growth in infancy and childhood and obesity in later life—a systematic review. *Obes Rev*. 2005;6:143-54. [PMID: 15836465]
118. Montori VM, Basu A, Erwin PJ, Velosa JA, Gabriel SE, Kudva YC. Posttransplantation diabetes: a systematic review of the literature. *Diabetes Care*. 2002;25:583-92. [PMID: 11874952]
119. Moreland JD, Richardson JA, Goldsmith CH, Clase CM. Muscle weakness and falls in older adults: a systematic review and meta-analysis. *J Am Geriatr Soc*. 2004;52:1121-9. [PMID: 15209650]
120. Needleman HL, Gatsonis CA. Low-level lead exposure and the IQ of children. A meta-analysis of modern studies. *JAMA*. 1990;263:673-8. [PMID: 2136923]
121. Newsome CA, Shiell AW, Fall CH, Phillips DI, Shier R, Law CM. Is birth weight related to later glucose and insulin metabolism?—a systematic review. *Diabet Med*. 2003;20:339-48. [PMID: 12752481]
122. O'Leary D, Paykel E, Todd C, Vardulaki K. Suicide in primary affective disorders revisited: a systematic review by treatment era. *J Clin Psychiatry*. 2001;62:804-11. [PMID: 11816870]
123. Oliver D, Daly F, Martin FC, McMurdo ME. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age Ageing*. 2004;33:122-30. [PMID: 14960426]
124. Pengel LH, Herbert RD, Maher CG, Refshauge KM. Acute low back pain: systematic review of its prognosis. *BMJ*. 2003;327:323. [PMID: 12907487]
125. Petticrew M, Bell R, Hunter D. Influence of psychological coping on survival and recurrence in people with cancer: systematic review. *BMJ*. 2002;325:1066. [PMID: 12424165]
126. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine*. 2002;27:E109-20. [PMID: 11880847]
127. Pocock SJ, Smith M, Baghurst P. Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. *BMJ*. 1994;309:1189-97. [PMID: 7987149]
128. Poobalan A, Aucott L, Smith WC, Avenell A, Jung R, Broom J, et al. Effects of weight loss in overweight/obese individuals and long-term lipid outcomes—a systematic review. *Obes Rev*. 2004;5:43-50. [PMID: 14969506]
129. Ramirez AJ, Westcombe AM, Burgess CC, Sutton S, Littlejohns P, Richards MA. Factors predicting delayed presentation of symptomatic breast cancer: a systematic review. *Lancet*. 1999;353:1127-31. [PMID: 10209975]
130. Reas DL, Schoemaker C, Zipfel S, Williamson DA. Prognostic value of duration of illness and early intervention in bulimia nervosa: a systematic review of the outcome literature. *Int J Eat Disord*. 2001;30:1-10. [PMID: 11439404]
131. Reid MC, Boutros NN, O'Connor PG, Cadariu A, Concato J. The health-related effects of alcohol use in older persons: a systematic review. *Subst Abuse*. 2002;23:149-64. [PMID: 12444348]
132. Rosenberg AL, Watts C. Patients readmitted to ICUs: a systematic review of risk factors and outcomes. *Chest*. 2000;118:492-502. [PMID: 10936146]
133. Ross SD, Estok RP, Frame D, Stone LR, Ludensky V, Levine CB. Disability and chronic fatigue syndrome: a focus on function. *Arch Intern Med*. 2004;164:1098-107. [PMID: 15159267]

134. Rubenstein JH, Laine L. Systematic review: the hepatotoxicity of non-steroidal anti-inflammatory drugs. *Aliment Pharmacol Ther.* 2004;20:373-80. [PMID: 15298630]
135. Rubin GJ, Hardy R, Hotopf M. A systematic review and meta-analysis of the incidence and severity of postoperative fatigue. *J Psychosom Res.* 2004;57:317-26. [PMID: 15507259]
136. Scannapieco FA, Bush RB, Paju S. Associations between periodontal disease and risk for nosocomial bacterial pneumonia and chronic obstructive pulmonary disease. A systematic review. *Ann Periodontol.* 2003;8:54-69. [PMID: 14971248]
137. Scannapieco FA, Bush RB, Paju S. Associations between periodontal disease and risk for atherosclerosis, cardiovascular disease, and stroke. A systematic review. *Ann Periodontol.* 2003;8:38-53. [PMID: 14971247]
138. Scannapieco FA, Bush RB, Paju S. Periodontal disease as a risk factor for adverse pregnancy outcomes. A systematic review. *Ann Periodontol.* 2003;8:70-8. [PMID: 14971249]
139. Schoemaker C. Does early intervention improve the prognosis in anorexia nervosa? A systematic review of the treatment-outcome literature. *Int J Eat Disord.* 1997;21:1-15. [PMID: 8986512]
140. Schoetzau A, Gehring U, Wichmann HE. Prospective cohort studies using hydrolysed formulas for allergy prevention in atopy-prone newborns: a systematic review. *Eur J Pediatr.* 2001;160:323-32. [PMID: 11421410]
141. Scholten-Peeters GG, Verhagen AP, Bekkering GE, van der Windt DA, Barnsley L, Oostendorp RA, et al. Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies. *Pain.* 2003;104:303-22. [PMID: 12855341]
142. Shekelle PG, Morton SC, Clark KA, Pathak M, Vickrey BG. Systematic review of risk factors for urinary tract infection in adults with spinal cord dysfunction. *J Spinal Cord Med.* 1999;22:258-72. [PMID: 10751130]
143. Sin DD, Wu L, Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest.* 2005;127:1952-9. [PMID: 15947307]
144. Spencer N, Logan S. Sudden unexpected death in infancy and socioeconomic status: a systematic review. *J Epidemiol Community Health.* 2004;58:366-73. [PMID: 15082732]
145. Steels E, Paesmans M, Berghmans T, Branle F, Lemaitre F, Mascaux C, et al. Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis. *Eur Respir J.* 2001;18:705-19. [PMID: 11716177]
146. Stock SR. Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. *Am J Ind Med.* 1991;19:87-107. [PMID: 1824910]
147. Stolwijk-Swüste JM, Beelen A, Lankhorst GJ, Nollet F. The course of functional status and muscle strength in patients with late-onset sequelae of poliomyelitis: a systematic review. *Arch Phys Med Rehabil.* 2005;86:1693-701. [PMID: 16084828]
148. Thacker SB, Stroup DF, Branche CM, Gilchrist J, Goodman RA, Porter Kelling E. Prevention of knee injuries in sports. A systematic review of the literature. *J Sports Med Phys Fitness.* 2003;43:165-79. [PMID: 12853898]
149. Thomas SL, Newell ML, Peckham CS, Ades AE, Hall AJ. A review of hepatitis C virus (HCV) vertical transmission: risks of transmission to infants born to mothers with and without HCV viraemia or human immunodeficiency virus infection. *Int J Epidemiol.* 1998;27:108-17. [PMID: 9563703]
150. Turner C, McClure R, Pirozzo S. Injury and risk-taking behavior—a systematic review. *Accid Anal Prev.* 2004;36:93-101. [PMID: 14572831]
151. Uzzan B, Nicolas P, Cucherat M, Perret GY. Microvessel density as a prognostic factor in women with breast cancer: a systematic review of the literature and meta-analysis. *Cancer Res.* 2004;64:2941-55. [PMID: 15126324]
152. van Dalen EC, van der Pal HJ, Bakker PJ, Caron HN, Kremer LC. Cumulative incidence and risk factors of mitoxantrone-induced cardiotoxicity in children: a systematic review. *Eur J Cancer.* 2004;40:643-52. [PMID: 15010064]
153. van Dam RM, Hu FB. Coffee consumption and risk of type 2 diabetes: a systematic review. *JAMA.* 2005;294:97-104. [PMID: 15998896]
154. van den Born BJ, Hulsman CA, Hoekstra JB, Schlingemann RO, van Montfrans GA. Value of routine funduscopy in patients with hypertension: systematic review. *BMJ.* 2005;331:73. [PMID: 16002881]
155. van der Windt DA, Thomas E, Pope DP, de Winter AF, Macfarlane GJ, Bouter LM, et al. Occupational risk factors for shoulder pain: a systematic review. *Occup Environ Med.* 2000;57:433-42. [PMID: 10854494]
156. van Dijk D, Keizer AM, Diephuis JC, Durand C, Vos LJ, Hijman R. Neurocognitive dysfunction after coronary artery bypass surgery: a systematic review. *J Thorac Cardiovasc Surg.* 2000;120:632-9. [PMID: 11003741]
157. van Dijk MA, Reitsma JB, Fischer JC, Sanders GT. Indications for requesting laboratory tests for concurrent diseases in patients with carpal tunnel syndrome: a systematic review. *Clin Chem.* 2003;49:1437-44. [PMID: 12928223]
158. Viganò A, Dorgan M, Buckingham J, Bruera E, Suarez-Almazor ME. Survival prediction in terminal cancer patients: a systematic review of the medical literature. *Palliat Med.* 2000;14:363-74. [PMID: 11064783]
159. Wang WH, Huang JQ, Zheng GF, Lam SK, Karlberg J, Wong BC. Non-steroidal anti-inflammatory drug use and the risk of gastric cancer: a systematic review and meta-analysis. *J Natl Cancer Inst.* 2003;95:1784-91. [PMID: 14652240]
160. Wasswa-Kintu S, Gan WQ, Man SF, Pare PD, Sin DD. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. *Thorax.* 2005;60:570-5. [PMID: 15994265]
161. Watine J, Miédougé M, Friedberg B. Carcinoembryonic antigen as an independent prognostic factor of recurrence and survival in patients resected for colorectal liver metastases: a systematic review. *Dis Colon Rectum.* 2001;44:1791-9. [PMID: 11742164]
162. Watine J. Laboratory variables as additional staging parameters in patients with small-cell lung carcinoma. A systematic review. *Clin Chem Lab Med.* 1999;37:931-8. [PMID: 10616746]
163. Watine J. Prognostic evaluation of primary non-small cell lung carcinoma patients using biological fluid variables. A systematic review. *Scand J Clin Lab Invest.* 2000;60:259-73. [PMID: 10943596]
164. Welch TP. Risk of giardiasis from consumption of wilderness water in North America: a systematic review of epidemiologic data. *Int J Infect Dis.* 2000;4:100-3. [PMID: 10737847]
165. Wojcicki JM. Socioeconomic status as a risk factor for HIV infection in women in East, Central and Southern Africa: a systematic review. *J Biosoc Sci.* 2005;37:1-36. [PMID: 15688569]
166. Wu O, Robertson L, Langhorne P, Twaddle S, Lowe GD, Clark P, et al. Oral contraceptives, hormone replacement therapy, thrombophilias and risk of venous thromboembolism: a systematic review. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) Study. *Thromb Haemost.* 2005;94:17-25. [PMID: 16113779]
167. Wu YW, Colford JM Jr. Chorioamnionitis as a risk factor for cerebral palsy: a meta-analysis. *JAMA.* 2000;284:1417-24. [PMID: 10989405]
168. Wulsin LR, Singal BM. Do depressive symptoms increase the risk for the onset of coronary disease? A systematic quantitative review. *Psychosom Med.* 2003;65:201-10. [PMID: 12651987]
169. Xu L, McElduff P, D'Este C, Attia J. Does dietary calcium have a protective effect on bone fractures in women? A meta-analysis of observational studies. *Br J Nutr.* 2004;91:625-34. [PMID: 15035690]
170. Zeegers MP, Tan FE, Dorant E, van Den Brandt PA. The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. *Cancer.* 2000;89:630-9. [PMID: 10931463]
171. Ariëns GA, van Mechelen W, Bongers PM, Bouter LM, van der Wal G. Physical risk factors for neck pain. *Scand J Work Environ Health.* 2000;26:7-19. [PMID: 10744172]
172. Almeida OP, Hulse GK, Lawrence D, Flicker L. Smoking as a risk factor for Alzheimer's disease: contrasting evidence from a systematic review of case-control and cohort studies. *Addiction.* 2002;97:15-28. [PMID: 11895267]
173. Bernal-Delgado E, Latour-Pérez J, Pradas-Arnal F, Gómez-López LI. The association between vasectomy and prostate cancer: a systematic review of the literature. *Fertil Steril.* 1998;70:191-200. [PMID: 9696205]
174. Hernández-Díaz S, Rodríguez LA. Association between nonsteroidal anti-inflammatory drugs and upper gastrointestinal tract bleeding/perforation: an overview of epidemiologic studies published in the 1990s. *Arch Intern Med.* 2000;160:2093-9. [PMID: 10904451]
175. Arenz S, Rückert R, Koletzko B, von Kries R. Breast-feeding and childhood obesity—a systematic review. *Int J Obes Relat Metab Disord.* 2004;28:1247-56. [PMID: 15314625]
176. Wortman PM. Judging research quality. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis.* New York: Russell Sage Foundation; 1994:97-109.

Current Author Addresses: Drs. Hayden and Côté: Institute for Work and Health, 481 University Avenue, Suite 800, 8th Floor, Toronto, Ontario M5G 2E9, Canada.

Dr. Bombardier: Toronto General Hospital, Eaton North Wing, 6th Floor, Room 231A, 200 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada.