Assessing Bias in Studies of Prognostic Factors

Article *in* Annals of Internal Medicine · February 2013

DOI: 10.7326/0003-4819-158-4-201302190-00009 · Source: PubMed

5 authors, including:

Danielle A van der Windt
Keele University
419 PUBLICATIONS **29,812** CITATIONS

SEE PROFILE

Jennifer L Cartwright
Dalhousie University
10 PUBLICATIONS **2,425** CITATIONS

SEE PROFILE

Pierre Côté
Ontario Tech University
412 PUBLICATIONS **22,182** CITATIONS

SEE PROFILE

Claire Bombardier
University of Toronto
618 PUBLICATIONS **97,214** CITATIONS

SEE PROFILE

Assessing Bias in Studies of Prognostic Factors

Jill A. Hayden, DC, PhD; Danielle A. van der Windt, PhD; Jennifer L. Cartwright, MSc; Pierre Côté, DC, PhD; and Claire Bombardier, MD

Previous work has identified 6 important areas to consider when evaluating validity and bias in studies of prognostic factors: participation, attrition, prognostic factor measurement, confounding measurement and account, outcome measurement, and analysis and reporting. This article describes the Quality In Prognosis Studies tool, which includes questions related to these areas that can inform judgments of risk of bias in prognostic research.

A working group comprising epidemiologists, statisticians, and clinicians developed the tool as they considered prognosis studies of low back pain. Forty-three groups reviewing studies addressing prognosis in other topic areas used the tool and provided feedback. Most reviewers (74%) reported that reaching consensus on judgments was easy. Median completion time per study was 20 minutes; interrater agreement (κ statistic) reported by 9 review teams varied from 0.56 to 0.82 (median, 0.75). Some reviewers reported challenges making judgments across prompting items, which were addressed by providing comprehensive guidance and examples. The refined Quality In Prognosis Studies tool may be useful to assess the risk of bias in studies of prognostic factors.

Ann Intern Med. 2013;158:280-286. www.annals.org
For author affiliations, see end of text.

Well-conducted prognostic research is important for clinical decision making. It informs patients about possible outcomes, identifies risk groups for stratified management, and helps target specific prognostic factors for modification (1). However, previous research shows many methodological shortcomings in the design and conduct of studies that address prognosis (2–4).

Critical appraisal of prognostic studies is essential to assess and identify biases sufficiently large to distort study results. A tool to guide such critical appraisal would help reviewers conducting systematic reviews and developing clinical practice guidelines, researchers conducting primary studies, and readers of such studies.

During assessment of risk of bias, 6 important domains should be considered when evaluating validity and bias in studies of prognostic factors: study participation, study attrition, prognostic factor measurement, confounding measurement and account, outcome measurement, and analysis and reporting (1). Researchers have used these recommendations to guide design and conduct of primary prognosis studies (5, 6) and as a guideline to improve reporting (6). In this article, we describe the refinement and use of the Quality In Prognosis Studies (QUIPS) tool to assess risk of bias in studies of prognostic factors.

Methods

Development of the QUIPS Tool

The **Figure** shows a schematic of the project. Fourteen working group members, including epidemiologists, statisticians, and clinicians, collaborated in tool development (7). The working group used an e-mail–based, modified Delphi approach (8) and nominal group techniques to refine prompting items for assessing bias domains and proposed ratings for the bias assessments as they considered prognosis studies of low back pain.

During an in-person workshop in 2006 that included working group members and other participants, a facilitator presented issues of agreement or dissent related to assessment of the bias domains. Through an iterative process of discussion and voting, workshop participants reached consensus on the wording of prompting items to guide ratings of high, moderate, or low risk of bias related to the 6 domains. These recommendations were formatted as a paper and an electronic tool and were used to assess risk of bias in studies included in a systematic review of prognostic factors in back pain (9). An overlapping group of 22 experts further discussed and refined the tool before and during a workshop in 2007.

Use of the Tool and Feedback

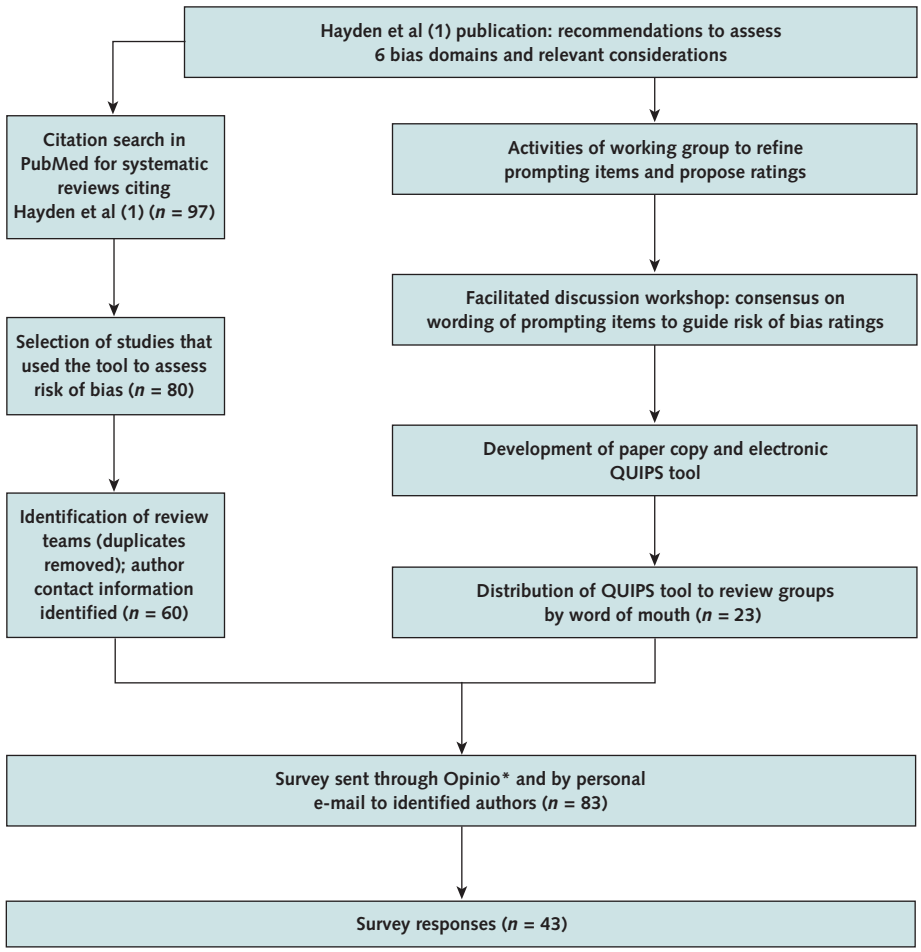
Since 2007, preliminary versions and subsequently a refined electronic version of the QUIPS tool were shared with and adapted by other research teams conducting systematic reviews of studies addressing prognosis, including review teams in rheumatology (10), cardiovascular disease (11, 12), and kidney disease (13, 14). We then used a structured Web-based survey to solicit feedback from 83 research teams that had used the QUIPS tool. Potential authors were identified by using a citation search in PubMed for the original 2006 QUIPS paper (1) and by reviewing personal communications that the primary investigator had received (**Figure**).

The survey was constructed using Opinio (ObjectPlanet, Oslo, Norway). We collected information on the characteristics of the systematic reviews that had used the QUIPS tool (such as topic area and review status), characteristics of the review teams (number of reviewers involved in the quality assessment process), how the tool was used (domains used, aspects of the tool used for quality assessment, and risk of bias judgments), its perceived ease of use (time to complete an assessment by using the tool and

See also:

Web-Only
Supplement

Figure. **Schematic of the project from 2006 through 2011 to develop and assess the QUIPS tool for assessing risk of bias in prognostic factor studies.**



We selected review teams for the survey if they conducted a prognosis systematic review, cited Hayden and colleagues (1) with reference to critical appraisal of included studies, and used a tool that sufficiently resembled the QUIPS tool (that is, included at least 4 of 6 domains of the QUIPS tool). QUIPS = Quality In Prognosis Studies.
* ObjectPlanet, Oslo, Norway.

problems encountered), and any suggested modifications. A copy of the complete survey is available on request.

Role of the Funding Source

There was no direct funding for this project.

Results

The QUIPS Tool

The **Table** summarizes the 6 bias domains, prompting items and considerations for each domain, and overall rating assessments. The **Supplement**, available at www.annals.org, shows the full version of the QUIPS tool.

The Study Participation domain addresses the representativeness of the study sample. It helps the assessor judge whether the study's reported association is a valid estimate of the true relationship between the prognostic factor and the outcome of interest in the source population. To make this judgment, the assessor considers the proportion of eligible persons who participate in the study, as well as descriptions of the source population, baseline study sample, sampling frame and recruitment, and inclusion and exclusion criteria. A study would be considered as having high risk of bias if the participation rate is low, the study sample has a very different age and sex distribution from the source population, or a very selective rather than consecutive sample of eligible patients was recruited. Conversely, studies with high participation of eligible and consecutively recruited patients who have characteristics similar to those in the source population would have low risk of bias.

The Study Attrition domain addresses whether participants with follow-up data represent persons enrolled in the study. It helps the assessor judge whether the reported

Table. **Summary of the Bias Domains, Prompting Items, and Ratings of the QUIPS Tool***

Variable	Bias Domains			
	1. Study Participation	**2. Study Attrition**	**3. Prognostic Factor Measurement**	**4. Outcome Measurement**
Optimal study or characteristics of unbiased study	The study sample adequately represents the population of interest	The study data available (i.e., participants not lost to follow-up) adequately represent the study sample	The PF is measured in a similar way for all participants	The outcome of interest is measured in a similar way for all participants
Prompting items and considerations†	a. Adequate participation in the study by eligible persons	a. Adequate response rate for study participants	a. A clear definition or description of the PF is provided	a. A clear definition of the outcome is provided
	b. Description of the source population or population of interest	b. Description of attempts to collect information on participants who dropped out	b. Method of PF measurement is adequately valid and reliable	b. Method of outcome measurement used is adequately valid and reliable
	c. Description of the baseline study sample	c. Reasons for loss to follow-up are provided	c. Continuous variables are reported or appropriate cut points are used	c. The method and setting of outcome measurement is the same for all study participants
	d. Adequate description of the sampling frame and recruitment	d. Adequate description of participants lost to follow-up	d. The method and setting of measurement of PF is the same for all study participants	
	e. Adequate description of the period and place of recruitment	e. There are no important differences between participants who completed the study and those who did not	e. Adequate proportion of the study sample has complete data for the PF	
	f. Adequate description of inclusion and exclusion criteria		f. Appropriate methods of imputation are used for missing PF data	
Ratings‡				
High risk of bias	The relationship between the PF and outcome is very likely to be different for participants and eligible nonparticipants	The relationship between the PF and outcome is very likely to be different for completing and noncompleting participants	The measurement of the PF is very likely to be different for different levels of the outcome of interest	The measurement of the outcome is very likely to be different related to the baseline level of the PF
Moderate risk of bias	The relationship between the PF and outcome may be different for participants and eligible nonparticipants	The relationship between the PF and outcome may be different for completing and noncompleting participants	The measurement of the PF may be different for different levels of the outcome of interest	The measurement of the outcome may be different related to the baseline level of the PF
Low risk of bias	The relationship between the PF and outcome is unlikely to be different for participants and eligible nonparticipants	The relationship between the PF and outcome is unlikely to be different for completing and noncompleting participants	The measurement of the PF is unlikely to be different for different levels of the outcome of interest	The measurement of the outcome is unlikely to be different related to the baseline level of the PF

PF = prognostic factor; QUIPS= Quality In Prognosis Studies.
* The **Supplement** (available at www.annals.org) shows the full QUIPS tool.
† Prompting items are to guide the user's judgment about risk of bias for each domain and are taken together to inform the overall judgment of potential bias and facilitate consensus among reviewers for each of the 6 domains. Some items may not be relevant to the specific study or the review research question; modification/clarification of the prompting items for the specific review question is encouraged.
‡ Each domain is rated as high, moderate, or low risk of bias considering the prompting items.

association between the prognostic factor and outcome is biased by the assessment of outcomes in a selected group of participants who completed the study. To make this judgment the assessor considers the study withdrawal rate (that is, whether many participants withdrew and whether there is a higher risk for systematic differences that may bias the prognostic factor association), information about why participants were lost to follow-up (that is, there is less concern if all persons provide random explanations), and observed differences in characteristics of persons lost to follow-up compared with participants who completed the study.

A study would be considered to have high risk of bias if it is probable that persons who completed the study differ from those lost to follow-up in a way that distorts the association between the prognostic factor and outcome. Conversely, studies with complete follow-up, or evidence of participants missing at random, have low risk of bias.

Table—Continued

Bias Domains	
5. Study Confounding	**6. Statistical Analysis and Reporting**
Important potential confounding factors are appropriately accounted for	The statistical analysis is appropriate, and all primary outcomes are reported
a. All important confounders are measured	a. Sufficient presentation of data to assess the adequacy of the analytic strategy
b. Clear definitions of the important confounders measured are provided	b. Strategy for model building is appropriate and is based on a conceptual framework or model
c. Measurement of all important confounders is adequately valid and reliable	c. The selected statistical model is adequate for the design of the study
d. The method and setting of confounding measurement are the same for all study participants	d. There is no selective reporting of results
e. Appropriate methods are used if imputation is used for missing confounder data	
f. Important potential confounders are accounted for in the study design	
g. Important potential confounders are accounted for in the analysis	
The observed effect of the PF on the outcome is very likely to be distorted by another factor related to PF and outcome	The reported results are very likely to be spurious or biased related to analysis or reporting
The observed effect of the PF on outcome may be distorted by another factor related to PF and outcome	The reported results may be spurious or biased related to analysis or reporting
The observed effect of the PF on outcome is unlikely to be distorted by another factor related to PF and outcome	The reported results are unlikely to be spurious or biased related to analysis or reporting

The Prognostic Factor Measurement domain addresses adequacy of prognostic factor measurement. It helps the assessor judge whether the study measured the prognostic factor in a similar, valid, and reliable way for all participants. To make this judgment, the assessor considers the clarity of the definition of the prognostic factor, evidence on the validity and reliability of the measurement approach, and the similarity of measurement and appropriate reporting of the prognostic factor for all participants. Information considered may include outside sources on measurement properties, blind or independent measurement, and limited reliance on recall.

A study would be considered to have low risk of bias if the prognostic factor is measured similarly for all participants and uses a valid, reliable measure. Conversely, studies that use an unreliable method to measure the prognostic factor or use different approaches for participants that result in systematic misclassification have high risk of bias.

The Outcome Measurement domain addresses the adequacy of outcome measurement. It helps the assessor judge whether the study measured the outcome in a similar, reliable, and valid way for all participants. To make this judgment, the assessor considers the clarity of outcome definition, evidence on the validity and reliability of the measurement, and similarity of measurement (that is, similar setting, method of measurement, and follow-up duration) for different levels of the prognostic factor. Information considered may include relevant outside sources on measurement properties, blind measurement, and confirmation of outcome with another valid and reliable test to support a judgment.

A study would have high risk of bias if there is likely to be differential measurement of outcome related to the extent of exposure to the prognostic factor; for example, if cardiovascular outcomes are assessed more extensively in smokers than in nonsmokers. A study would be considered to have low risk of bias if the outcome is measured similarly for all participants and uses a valid, reliable measure.

The Study Confounding domain addresses potential confounding factors. It helps the assessor judge whether another factor may explain the study's reported association. To make this judgment, the assessor considers the validity, reliability, and similarity of measurement of potential confounders (defined a priori) for all participants and whether all important confounding factors are accounted for in the study design or analysis.

A study would have high risk of bias if another factor related to both the prognostic factor and the outcome is likely to explain the effect of the prognostic factor. Conversely, studies with adequate measurement of important potential confounding variables and inclusion of these variables in a prespecified multivariable analysis have low risk of bias.

The Statistical Analysis and Reporting domain addresses the appropriateness of the study's statistical analysis and completeness of reporting. It helps the assessor judge whether results are likely to be spurious or biased because of analysis or reporting. To make this judgment, the assessor considers the data presented to determine the adequacy of the analytic strategy and model-building process and investigates concerns about selective reporting. Selective reporting is an important issue in prognostic factor reviews because studies commonly report only factors positively associated with outcomes. A study would be considered to have low risk of bias if the statistical analysis is appropriate for the data, statistical assumptions are satisfied, and all primary outcomes are reported.

Using the Tool

For each of the 6 domains in the QUIPS tool, responses to the prompting items are taken together to inform the judgment of risk of bias. Information and methodological comments supporting the item assessment should be recorded (cited directly from the study publication). Judgments should be made with consensus among at least 2 assessors. Some items may not be relevant to the specific study or the review question and may be skipped or omitted. For example, if a study has a 100% response rate, the prompting items in the Study Attrition domain related to collection of information on participants who dropped out of the study, reasons for loss to follow-up, and description and comparison of key characteristics of participants lost to follow-up with study completers are not relevant.

To grade the tool, each of the 6 potential bias domains is rated as having high, moderate, or low risk of bias. For example, with respect to the Study Attrition domain, study A reported an 80% response rate (20% of the study sample lost to follow-up); the authors tried to determine reasons for noncompletion, collected and presented information about key characteristics of those lost to follow-up, and found no differences between completers and noncompleters on important characteristics and outcomes. This study was rated as having low risk of bias due to study attrition. Study B, however, would be judged as having high risk of bias due to attrition with the same 80% response rate if important systematic differences existed between participants who did and those who did not complete the study. Finally, study C would be judged as having low Risk of Bias due to attrition with only the information that 99% of a large study sample completed outcome assessment.

Assessing the overall risk of bias in each study may also be useful. To judge overall risk, one could describe studies with a low risk of bias as those in which all, or the most important (as determined a priori), of the 6 important bias domains are rated as having low risk of bias. We recommend use of sensitivity analyses to explore the effect of the selected definition. In line with the Cochrane Risk of Bias tool for intervention studies (15) and the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool for diagnostic studies (16), we recommend against the use of a summated score for overall study quality.

Feedback From Reviewers

Forty-three of the 83 review authors invited to provide feedback on the QUIPS tool did so (**Figure**). The reviews came from diverse topic areas, including musculoskeletal disorders (13 of 43 review teams), obstetrics and pediatrics (7 of 43), heart or vascular disease (6 of 43), and cancer (4 of 43). Most focused on prognostic factors (28 of 43 reviews), although some examined overall prognosis (6 of 43), risk prediction models (9 of 43), or differential treatment effect by prognostic factors (2 of 43).

Appendix Table 1 (available at www.annals.org) shows the experiences of the researchers who used the QUIPS tool. Most review teams had 2 reviewers independently complete risk of bias assessments and used consensus processes to resolve disagreements. Most review teams (28 of 38) reported that the process of reaching consensus on assessments was "easy." Interrater agreement, reported as percentage of agreement by 9 review teams (10, 17–24) on 205 studies (reported in peer-reviewed publications or by personal communication), varied between 70% and 89.5% (median, 83.5%).

The κ statistic for independent rating of QUIPS items, reported by 9 review teams (10, 19, 23, 25–30) on 159 studies, varied from 0.56 to 0.82 (median, 0.75). One review team (31 studies) (25) reported interrater agreement scores for individual bias domains: study participation ($\kappa = 0.73$), study attrition ($\kappa = 1.0$), prognostic factor measurement ($\kappa = 1.0$), confounding measurement and account ($\kappa = 0.4$), outcome measurement ($\kappa = 0.73$), and analysis and reporting ($\kappa = 0.73$).

Review teams reported that using the QUIPS tool took a median of 20 minutes per study; 5 reviewers reported that it took their team longer than 1 hour per study. Many review teams included members with specific training or education to complete the assessments.

Most review teams (32 of 42) used versions of the QUIPS tool that were developed from recommendations in Hayden and colleagues' article (1). Ten review teams had access to the refined electronic QUIPS tool. One team described combining the QUIPS recommendations with the items from the Reporting Recommendations for Tumour Marker Prognostic Studies reporting guidelines (31), and 2 review groups referred to versions from other authors (for example, the National Institute for Health and Clinical Excellence guideline manual [32]). Fifteen groups did not judge risk of bias for the 6 domains but rather rated only the prompting items. Approximately half of the reviewers (15 of 34) reported using a count or an algorithm of the prompting items, and half (16 of 34) used judgment considering prompting items to rate the domain and overall risk of bias (**Appendix Table 2**, available at www.annals.org).

The results of the risk of bias assessments were presented and used in various ways. The most common approaches were to present individual prompting item ratings for each included study and additionally report an assessment of overall study quality. Two reviewers presented no critical appraisal results in their reviews.

Although feedback was positive, some reviewers reported challenges. Two review teams reported that they had difficulty making judgments across multiple prompting items, and 2 review groups commented that poor reporting in their included studies made judgment difficult. Seventeen review teams reported that they had advanced epidemiologic training for assessors. Seven review teams, using the tool for types of prognosis reviews other than

prognostic factor reviews, commented that they modified the tool by adding items or removing unnecessary items or domains.

Discussion

The QUIPS tool supports a systematic appraisal of bias in studies of prognostic factors. It is based on recommendations from a comprehensive review of quality assessment in prognosis systematic reviews (1) and is informed by basic epidemiologic principles. Independently developed and modified versions of the tool have been successfully used by several research groups, with moderate to substantial interrater reliability.

We previously found that quality assessment in prognosis systematic reviews is inconsistent and often incomplete (1). A recent review of quality assessment in chronic disease epidemiology systematic reviews similarly found that only 55% of included reviews reported quality assessment (33). Sanderson and associates (34) reviewed published tools that assess risk of bias in observational epidemiology studies. Similar to our original review of quality assessment tools used in prognosis systematic reviews, they reported a lack of suitable tools (34). The QUIPS tool that we developed fills this gap and includes a comprehensive set of prompting items with clear suggestions for operationalization and grading.

Some review groups participating in this study commented on the need to modify and refine the prompting items and eliminate some overlap of items. We encourage operationalization of the tool for specific purposes, including specifying key characteristics (for example, potential confounders), omitting any irrelevant prompting items, and adding new items where needed. Clear specification of the tool items will probably increase interrater agreement. For systematic reviews, operationalization of the tool should be done a priori and authors should make their application of the tool accessible to readers of their published article.

The QUIPS tool was designed to assess prognostic factor studies; however, it can provide a starting point for development or refinement of quality assessment tools for other types of prognostic studies. For example, it may be modified to assess studies of overall prognosis (such as Moulaert and coworkers' systematic review [18]) by omitting domains related to prognostic factor measurement and confounding, along with slight adjustments to the prompting questions for the analysis domain.

Several review groups using the QUIPS tool reported counting prompting items as a scale. We recommend the assessment of prompting items to guide judgment of the 6 bias domains rather than using them as a scale. This approach involves balancing information about competing design or conduct features and is more transparent. However, we acknowledge that such a consensus-based judgment of potential bias is more challenging and requires assessors to be knowledgeable of epidemiologic methods. Online training tools and examples using the QUIPS tool should be developed to support training needs.

Our study has limitations. The group of experts who developed the tool were from a single topic area, potentially limiting generalizability. Furthermore, participants in our retrospective survey about the tool and our reported reliability scores were from a selected group of interested systematic reviewers. Our users probably have more advanced training and may overestimate usability and reliability scores for the wider population of potential users.

Future studies should further evaluate the QUIPS tool by using a prospective study design. Reliability testing should be done on a larger, more representative set of studies and tool users, including assessing reliability of individual domain ratings, as well as consensus ratings between groups. Exploring the effect of study-level factors on reliability of bias appraisal by using the QUIPS tool will also help identify potential problem areas in need of further guidance (35).

The relationship between domain ratings and prognostic factor associations to provide empirical evidence of design-related bias (that is, evidence of over- or underestimation of prognostic factor associations with judgments of increased bias related to each of the domains) needs to be examined. Our previous evaluation of systematic reviews of prognostic factors (1) found limited investigation of the association between study design characteristics and effect estimate (42 of 163 reviews reported), and findings were inconsistent for specific biases. Assessment of potential biases in prognosis studies included in systematic reviews by using a domain-based approach will facilitate future meta-epidemiologic studies to determine the effect of design-related biases.

Assessment of potential biases is particularly challenging in observational studies that are designed to investigate prognostic factors. The refined QUIPS tool is useful and reliable for systematic reviewers, study authors, and readers to guide comprehensive assessment of 6 bias domains in studies of prognostic factors.

From Dalhousie University, Halifax, Nova Scotia, Canada; Arthritis Research UK Primary Care Centre, Primary Care and Health Sciences, Keele University, Staffordshire, United Kingdom; University of Ontario Institute of Technology, Oshawa, Ontario, Canada; and University of Toronto and Institute for Work & Health, Toronto, Ontario, Canada.

Acknowledgment: The authors thank the QUIPS-Low Back Pain Working Group members (2006 and 2007) for their important contributions. They also thank the prognosis systematic review authors who completed their survey and review authors who responded to their additional questions and requests for data, including Amika Singh, James Chalmers, Roger Chou, Fiona Clay, Hanneke Creemers, Lotte Dyhrberg O'Neill, Jan Hartvigsen, Ross Iles, David Jimenez, Sindhu Johnson, Bindee Kuriya, Jolanda Luime, Veronique Moulaert, Tinca Polderman, Cara Wasywich, Stephen Wilton, Susan Woolfenden, Lexie Wright, and Christina Wyatt.

Financial Support: Dr. Hayden received infrastructure funding through the Nova Scotia Cochrane Resource Centre provided by the Nova Scotia Health Research Foundation and holds a Research Professorship in Epidemiology funded by the Canadian Chiropractic Research Foundation and Dalhousie University. Dr. van der Windt is a member of the Prognosis Research Strategy Initiative Medical Research Council, Prognosis Research Strategy Initiative Partnership (G0902393/99558).

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M12-1871.

Requests for Single Reprints: Jill A. Hayden, DC, PhD, Department of Community Health & Epidemiology, Dalhousie University, 5790 University Avenue, Room 222, Halifax, Nova Scotia B3H 1V7, Canada; e-mail, jhayden@dal.ca.

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

References

1. **Hayden JA, Côté P, Bombardier C.** Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144:427-37. [PMID: 16549855]
2. **Hemingway H, Riley RD, Altman DG.** Ten steps towards improving prognosis research. BMJ. 2009;339:b4184. [PMID: 20042483]
3. **Hayden JA, Chou R, Hogg-Johnson S, Bombardier C.** Systematic reviews of low back pain prognosis had variable methods and results: guidance for future prognosis reviews. J Clin Epidemiol. 2009;62:781-796.e1. [PMID: 19136234]
4. **Riley RD, Sauerbrei W, Altman DG.** Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. Br J Cancer. 2009;100:1219-29. [PMID: 19367280]
5. **Kamper SJ, Hancock MJ, Maher CG.** Optimal designs for prediction studies of whiplash. Spine (Phila Pa 1976). 2011;36:S268-74. [PMID: 22020594]
6. **Hemingway H, Philipson P, Chen R, Fitzpatrick NK, Damant J, Shipley M, et al.** Evaluating the quality of research into a single prognostic biomarker: a systematic review and meta-analysis of 83 studies of C-reactive protein in stable coronary artery disease. PLoS Med. 2010;7:e1000286. [PMID: 20532236]
7. **Hayden JA, Côté P, Steenstra IA, Bombardier C; QUIPS-LBP Working Group.** Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. J Clin Epidemiol. 2008;61:552-60. [PMID: 18471659]
8. **Jones J, Hunter D.** Consensus methods for medical and health services research. BMJ. 1995;311:376-80. [PMID: 7640549]
9. **Hayden JA.** Methodological Issues in Systematic Reviews of Prognosis and Prognostic Factors: Low Back Pain. Toronto: Univ Toronto; 2007.
10. **Chapple CM, Nicholson H, Baxter GD, Abbott JH.** Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. Arthritis Care Res (Hoboken). 2011;63:1115-25. [PMID: 21560257]
11. **Pickett CA, Jackson JL, Hemann BA, Atwood JE.** Carotid bruits as a prognostic indicator of cardiovascular death and myocardial infarction: a meta-analysis. Lancet. 2008;371:1587-94. [PMID: 18468542]
12. **Pickett CA, Jackson JL, Hemann BA, Atwood JE.** Carotid bruits and cerebrovascular disease risk: a meta-analysis. Stroke. 2010;41:2295-302. [PMID: 20724720]
13. **Palmer SC, Hayen A, Macaskill P, Pellegrini F, Craig JC, Elder GJ, et al.** Serum levels of phosphorus, parathyroid hormone, and calcium and risks of death and cardiovascular disease in individuals with chronic kidney disease: a systematic review and meta-analysis. JAMA. 2011;305:1119-27. [PMID: 21406649]
14. **Mathew A, Devereaux PJ, O'Hare A, Tonelli M, Thiessen-Philbrook H, Nevis IF, et al.** Chronic kidney disease and postoperative mortality: a systematic review and meta-analysis. Kidney Int. 2008;73:1069-81. [PMID: 18288098]
15. **Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al; Cochrane Bias Methods Group.** The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928. [PMID: 22008217]
16. **Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al; QUADAS-2 Group.** QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-36. [PMID: 22007046]
17. **Singh AS, Mulder C, Twisk JW, van Mechelen W, Chinapaw MJ.** Tracking of childhood overweight into adulthood: a systematic review of the literature. Obes Rev. 2008;9:474-88. [PMID: 18331423]
18. **Moulaert VR, Verbunt JA, van Heugten CM, Wade DT.** Cognitive impairments in survivors of out-of-hospital cardiac arrest: a systematic review. Resuscitation. 2009;80:297-305. [PMID: 19117659]
19. **van Drongelen A, Boot CR, Merkus SL, Smid T, van der Beek AJ.** The effects of shift work on body weight change—a systematic review of longitudinal studies. Scand J Work Environ Health. 2011;37:263-75. [PMID: 21243319]
20. **Clay FJ, Newstead SV, McClure RJ.** A systematic review of early prognostic factors for return to work following acute orthopaedic trauma. Injury. 2010;41:787-803. [PMID: 20435304]
21. **Nijrolder I, van der Horst H, van der Windt D.** Prognosis of fatigue. A systematic review. J Psychosom Res. 2008;64:335-49. [PMID: 18374732]
22. **Proper KI, Singh AS, van Mechelen W, Chinapaw MJ.** Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. Am J Prev Med. 2011;40:174-82. [PMID: 21238866]
23. **Spee LA, Madderom MB, Pijpers M, van Leeuwen Y, Berger MY.** Association between helicobacter pylori and gastrointestinal symptoms in children. Pediatrics. 2010;125:e651-69. [PMID: 20156901]
24. **van Duijvenbode DC, Hoozemans MJ, van Poppel MN, Proper KI.** The relationship between overweight and obesity, and sick leave: a systematic review. Int J Obes (Lond). 2009;33:807-16. [PMID: 19528969]
25. **Jiménez D, Uresandi F, Otero R, Lobo JL, Monreal M, Martí D, et al.** Troponin-based risk stratification of patients with acute nonmassive pulmonary embolism: systematic review and metaanalysis. Chest. 2009;136:974-82. [PMID: 19465511]
26. **Wright AA, Cook C, Abbott JH.** Variables associated with the progression of hip osteoarthritis: a systematic review. Arthritis Rheum. 2009;61:925-36. [PMID: 19565541]
27. **Elshout G, Monteny M, van der Wouden JC, Koes BW, Berger MY.** Duration of fever and serious bacterial infections in children: a systematic review. BMC Fam Pract. 2011;12:33. [PMID: 21575193]
28. **Gieteling MJ, Bierma-Zeinstra SM, Lisman-van Leeuwen Y, Passchier J, Berger MY.** Prognostic factors for persistence of chronic abdominal pain in children. J Pediatr Gastroenterol Nutr. 2011;52:154-61. [PMID: 21057328]
29. **Jeejeebhoy FM, Zelop CM, Windrim R, Carvalho JC, Dorian P, Morrison LJ.** Management of cardiac arrest in pregnancy: a systematic review. Resuscitation. 2011;82:801-9. [PMID: 21549495]
30. **Johnson SR, Swiston JR, Swinton JR, Granton JT.** Prognostic factors for survival in scleroderma associated pulmonary arterial hypertension. J Rheumatol. 2008;35:1584-90. [PMID: 18597400]
31. **McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM; Statistics Subcommittee of NCI-EORTC Working Group on Cancer Diagnostics.** REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat. 2006;100:229-35. [PMID: 16932852]
32. **National Institute for Health and Clinical Excellence.** Appendix J: methodology checklist: prognostic studies. In: The Guidelines Manual, London: National Institute for Health and Clinical Excellence; 2009:218-22.
33. **Shamliyan T, Kane RL, Jansen S.** Systematic reviews synthesized evidence without consistent quality assessment of primary studies examining epidemiology of chronic diseases. J Clin Epidemiol. 2012;65:610-8. [PMID: 22424987]
34. **Sanderson S, Tatt ID, Higgins JP.** Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol. 2007;36:666-76. [PMID: 17470488]
35. **Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, et al.** Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville, MD: Agency for Healthcare Research and Quality; 2012. Accessed at www.ncbi.nlm.nih.gov/books/NBK92293/ on 20 November 2012.

Current Author Addresses: Dr. Hayden: Department of Community Health & Epidemiology, Dalhousie University, 5790 University Avenue, Room 222, Halifax, Nova Scotia B3H 1V7, Canada.
Dr. van der Windt: Arthritis Research UK Primary Care Centre, Primary Care Sciences, Keele University, Staffordshire ST5 5BG, United Kingdom.
Ms. Cartwright: Department of Community Health & Epidemiology, Dalhousie University, 5790 University Avenue, Room 228, Halifax, Nova Scotia B3H 1V7, Canada.
Dr. Côté: Faculty of Health Sciences, University of Ontario Institute of Technology, 2000 Simcoe Street North, Oshawa, Ontario L1H 7K4, Canada.
Dr. Bombardier: Toronto General Hospital, Eaton North Wing, 6th Floor, Room 231A, 200 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada.

Author Contributions: Conception and design: J.A. Hayden, D.A. van der Windt, P. Côté.
Analysis and interpretation of the data: J.A. Hayden, D.A. van der Windt, J.L. Cartwright, P. Côté, C. Bombardier.
Drafting of the article: J.A. Hayden, J.L. Cartwright, P. Côté.
Critical revision of the article for important intellectual content: J.A. Hayden, D.A. van der Windt, J.L. Cartwright, P. Côté, C. Bombardier.
Final approval of the article: J.A. Hayden, D.A. van der Windt, J.L. Cartwright, P. Côté, C. Bombardier.
Collection and assembly of data: J.A. Hayden, D.A. van der Windt, J.L. Cartwright.

Appendix Table 1. **Description of Experience of Review Teams Conducting Risk of Bias Assessment by Using the QUIPS Tool***

Characteristic of Critical Appraisal	Review Teams, *n*†
Number of reviewers involved in conducting the critical appraisal	
1	3
2	31
3 or 4	7
Process used for critical appraisal	
Single reviewer	2
Single reviewer with checking by a second reviewer	3
Independent evaluation by 2 reviewers with consensus	33
Independent evaluation by >2 reviewers with consensus	3
Other	1
Ease of reaching consensus on assessments	
Very easy	6
Easy	22
Neutral	7
Hard	3
Time to complete critical appraisal of each study	
Median time (range)	20 (5–90) min
>10 min	36
>20 min	23
>1 h	5
Training or education to complete the critical appraisal	
No	24
Yes	17

QUIPS= Quality In Prognosis Studies.
* Total number of review teams is 43.
† Where multiple choices were possible or questions have been skipped without providing an answer, the number of review teams may not always sum to 43.

Appendix Table 2. **Description of How the QUIPS Tool Was Used by Review Teams***

Question	Review Teams, *n*†
Number of QUIPS potential bias domains assessed	
All 6	29
5	9
4	4
QUIPS bias domains assessed	
Study participation	42
Study attrition	37
Prognostic factor measurement	41
Outcome measurement	41
Study confounding	35
Statistical analysis and reporting	39
How prompting items were used	
All prompting items were scored	24
Prompting items were used to guide judgments only	15
Other	1
How ratings of risk of bias for each domain were determined	
Count of items satisfied/not satisfied or algorithm to combine items	15
Overall judgment	16
Other	3
How the overall risk of bias of each study was rated	
Count or score of individual prompting items	13
Count or score of risk of bias domain assessments	7
Overall judgment	13
Overall quality of each study was not assessed	9
Presentation of critical appraisal results for studies included in review	
Reported ratings for individual items for each included study	20
Reported each risk of bias domain assessment for each included study	9
Reported an assessment of quality for each included study	20
Reported an overall assessment of quality across all included studies	12
No presentation of critical appraisal results for included studies	2
Use of the results of the critical appraisal in synthesizing review evidence	
Described the results of the quality assessment for all studies	24
Used the quality assessment items or score as inclusion/exclusion criteria	3
Used a quality score to define the level of study quality or to rank studies	19
Tested the association of potential biases and study results‡	5
Not used in synthesis	6

QUIPS= Quality In Prognosis Studies.
* Total number of review teams is 43.
† Where multiple choices were possible or questions have been skipped without providing an answer, the number of review teams may not always sum to 43.
‡ Using subgroup or metaregression analyses.