DAVID M. WALTON, PT, MSc[1] • JASON PRETTY, BA[2]
JOY C. MACDERMID, PT, PhD[3] • ROBERT W. TEASELL, MD, FRCPC[4]

Risk Factors for Persistent Problems Following Whiplash Injury: Results of a Systematic Review and Meta-analysis

• **STUDY DESIGN:** Systematic review and meta-analysis.

• **BACKGROUND:** Whiplash-associated disorder (WAD) is the most common reported injury following motor vehicle accident. Evidence for prognosis and intervention are difficult to interpret due to differences in inception times, outcomes used, and sample heterogeneity.

• **METHODS:** An extensive literature search was conducted to identify published studies of prognosis following whiplash. Rigorous inclusion criteria were applied to allow for meaningful results to be drawn. Data were extracted, transformed where necessary, and pooled to allow estimation of the odds ratio for any factor with at least 3 data points in the literature.

• **RESULTS:** From 11 cohorts (n = 3193), 25 factors were identified with at least 3 data points in the existing literature. Of these, 9 were found to be significant predictors based on the odds ratio and confidence limits: no postsecondary education, female gender, history of previous neck pain, baseline neck pain intensity greater than 55/100, presence of neck pain at baseline, presence of headache at baseline, catastrophizing, WAD grade 2 or 3, and no seat belt in use at time of collision. Neck pain intensity, WAD grade, headache, and no postsecondary education were robust to publication bias.

• **CONCLUSIONS:** Using a rigorous process for the identification and extraction of data from a homogenous subset of the prognostic WAD literature, we were able to identify several factors for which information is easy to collect clinically and could provide clinicians with a good sense of prognosis following whiplash injury.

• **LEVEL OF EVIDENCE:** Prognosis, level 1a.
J Orthop Sports Phys Ther 2009;39(5):334-350. doi:10.2519/jospt.2009.2765

• **KEYWORDS:** *cervical spine, neck, prognosis, WAD, whiplash-associated disorder*

Whiplash following motor vehicle collision is a condition with substantial social and, in many cases, personal cost. Whiplash is defined as an acceleration-deceleration mechanism of energy transfer to the neck, resulting from rear-end or side-impact collisions but also from diving or other mishaps.[31] Spitzer et al[31] coined the term "whiplash-associated disorder" (WAD) to describe the spectrum of signs and symptoms that may be experienced as a result of a whiplash injury. The annual incidence of WAD has been reported to range from 70 to 328 per 100 000 population in Canada[31] and the United States.[24] Insurance data report the indirect costs in Canada in 1995 to be estimated at approximately $2500 per case.[31] Patient history alone can establish that a whiplash injury has occurred. However, a sound method for determining the prognosis is more elusive. Findings of persistent problems in 12%[18] to 84%[20] of patients at 12-month follow-up suggest that protracted pain or disability following WAD is a common problem.

Since the Quebec Task Force (QTF)[31] found that there were few, if any, methodologically sound observational studies on which to base recommendations for prognosis, there has been an increase in cohort studies aimed at identifying risk factors for persistent WAD. Inconsistencies in time from injury to baseline data collection, time to follow-up, and outcomes have made it difficult to synthesize the literature in this area. Two previous attempts have been made since the QTF publication.[8,29] Both reviews included articles of various designs and methodologies, and both took the form of systematic review. The findings suggest that there is strong evidence for high baseline pain as a predictive factor, but provide little concrete information on the extent of risk or an estimate of clinical impact.

The purpose of this review and meta-analysis was to statistically synthesize the

[1]Graduate student, School of Physical Therapy, The University of Western Ontario, London, ON, Canada. [2]Research Assistant, Lawson Health Research Institute, Department of Physical Medicine and Rehabilitation, Parkwood Hospital, London, ON, Canada. [3]Co-Director, Clinical Research Lab, Hand and Upper Limb Center, St Joseph's Health Care, London, ON, Canada. [4]Chair-Chief, Department of Physical Medicine and Rehabilitation, Parkwood Hospital, London, ON, Canada; Professor, Schulich School of Medicine, The University of Western Ontario, London, ON, Canada. David Walton is supported by a Doctoral Fellowship through the Canadian Institutes of Health Research. Joy MacDermid is supported by a New Researcher Award through the Canadian Institutes of Health Research. Address correspondence to David M. Walton, School of Physical Therapy, Rm EC1443, 1201 Western Rd, The University of Western Ontario, London, ON, Canada, N6G 1H1. E-mail: dwalton5@uwo.ca

findings from a more homogenous subset of prognostic studies to determine an estimate of the size and strength of the odds ratio for potential predictors of persistent WAD-related pain or disability, when collected within the first 3 weeks following injury. This information will allow for numerical comparisons of the relative increase in the odds that a patient will develop persistent pain or disability, based on the presence or absence of the factors identified herein.

METHODS

Search Strategy

An extensive electronic literature search was conducted of 4 international databases of scientific literature (Medline, CINAHL, Embase, and PsycInfo) from 1995 to May 2007. We chose to limit our search to articles published in or after 1995 because this is the year that the QTF monograph was released, providing the first standardized definition of WAD. It is also notable that the QTF found no published articles of adequate methodological rigor on the topic published up to 1993.[31]

The search terms used were "whiplash" or "traffic accident," paired with "prognosis," "prospective" or "cohort," and "neck." A secondary search was conducted through manual examination of reference lists of relevant literature identified in the first search. In total, 447 articles were reviewed for relevance to the topic (**FIGURE 1**).

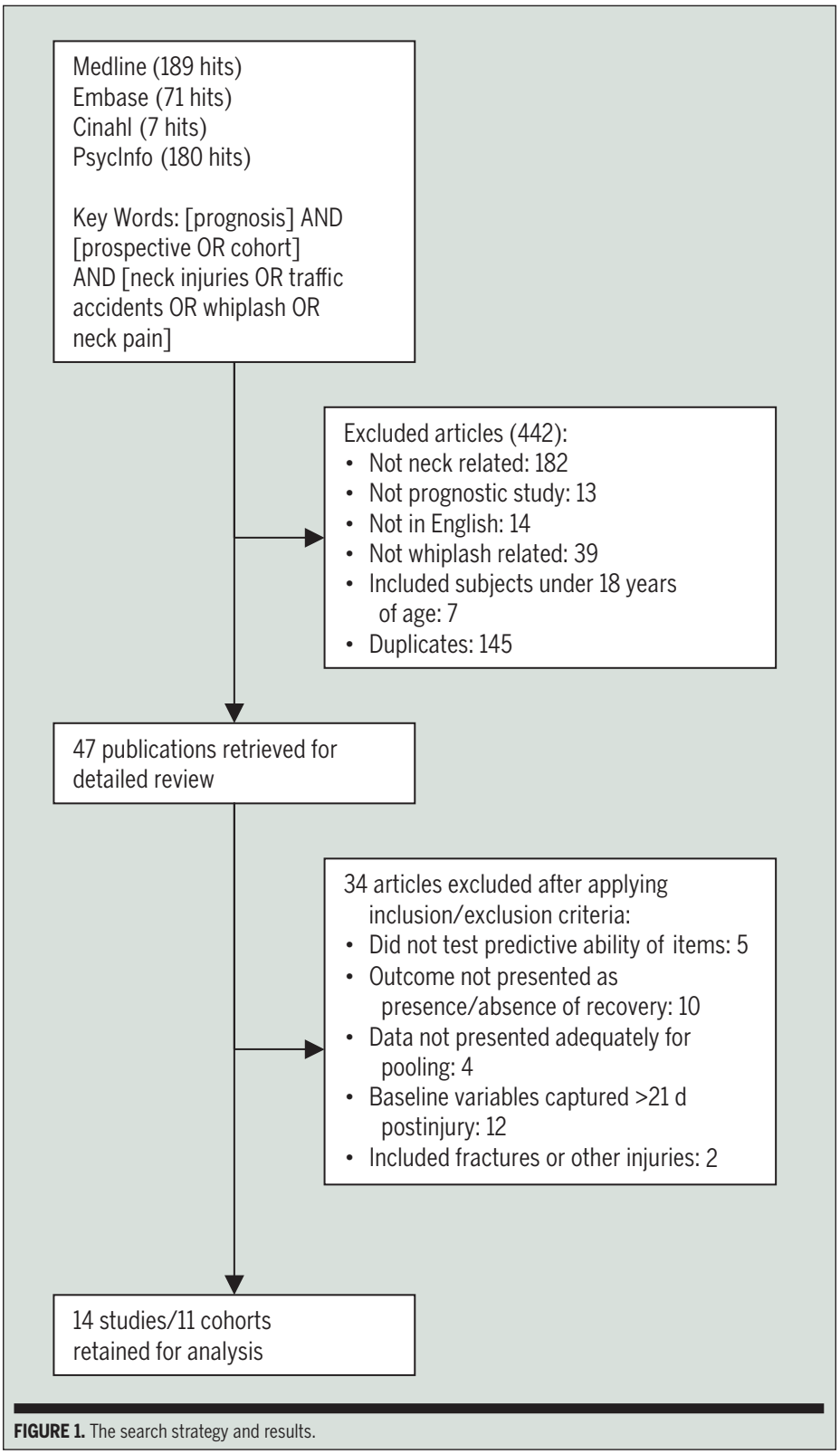


FIGURE 1. The search strategy and results.

Inclusion Criteria

Articles were included for further review if they met the following criteria: (1) the authors performed a prospective evaluation of 1 or more clinical risk factors for chronicity, (2) the inception (inclusion in study) and all baseline data collection occurred within 3 weeks of a motor vehicle accident for all subjects, (3) subjects were followed for at least 6 months postinjury to determine the presence of ongoing WAD-related symptoms and/or disability, (4) subjects with serious injuries, including fracture of the skull or vertebral column, paralysis, or significant brain injury, were excluded, and (5) all subjects were 18 years of age or older. Studies were also required to present sufficient data to allow for the calculation of effect size

TABLE 1 Descriptions of Cohorts Included in the Meta-analysis

Cohort, Authors	Sampling Frame	Time to Follow-up	Outcome Collected	Sample Size	Study Quality*
1. Atherton et al (2006)	Manchester, UK emergency department	12 mo	Report of persistent neck pain at 1, 3, and 12 mo postinjury	480	25
Predictors followed: General health prior to collision, lifetime experience of neck pain, preinjury widespread body pain, general practitioner consultation in year prior to collision, job satisfaction, boring or monotonous work, fast-paced or hectic work, stressful work, able to make decision at work, ability to learn new things at work, dissatisfaction with support from boss and colleagues, in own vehicle, collision severity, speed of own vehicle, speed of other vehicle, direction of collision, awareness of collision, position in vehicle, headrest type, seat belt use, airbag, initial injury severity (subjective), initial neck disability index score, number of other whiplash-associated disorders (WADs) symptoms, WAD classification, cervical bony tenderness, presence of neurological signs or symptoms, limited range of motion.				
2. Berglund et al (2006)	Sweden insurance claims	6, 12, 24 mo	Neck pain intensity on numeric rating scale (low, moderate, high); Disability Rating Index (low, moderate, high); Hospital Anxiety and Depression Scale (present or absent)	1391	27
Predictors followed: Gender, age, position in vehicle, individual disposable income, education, type of motor vehicle crash (MVC), awareness of collision, use of head rest, use of seat belt, head position at impact, car seat broken, initial neck pain intensity, initial headache, self-reported whiplash severity (neck complaints only), neck complaints and perceived reduced neck movements, neck complaints and perceived reduced neck movements and numbness or pain in arms or hands.†					
3. Borchgrevink et al (1997)	Norway emergency department	6 mo	Headache, neck pain, or neck stiffness every day or constantly, using 4-point ordinal scale	88	15
Predictors followed: Age, gender, employment status, family income, collision parameters, personality as measured by MCMI-1 personality inventory.					
4. Hartling et al (2000, 2001, 2002)	Ontario emergency department	6 mo‡	Ongoing WAD, based on frequency and intensity of symptoms	334	21
Predictors followed: Symptoms before accident, location of initial pain, other symptoms (headache, neck stiffness, upper extremity numbness or weakness, dizziness, nausea/vomiting, vision disturbances, difficulty sleeping, fatigue, anxiety, difficulty concentrating, depression, total number of initial symptoms), initial severity and frequency of symptoms, position in vehicle, location of accident, road conditions, prepared for crash, head position, vehicle pushed forward, vehicle hit another in front, seat belt in use, head restraint, head snapped back, head went over headrest, transmission, brake light in rear window, size of vehicle that hit, posted speed limit, gender, age, height, body mass index, missed time from work, amount of time missed from work, changed work activities, changed leisure activities, WAD category.				
5. Hendriks et al (2005)	The Netherlands primary care	12 mo	Recovered: pain intensity <30 mm or work activities >78 mm and not on pain medication	119	27
Predictors followed: Age, gender, no postsecondary education, marital status, private insurance, employment status, direction of impact, location in vehicle, seat belt use, unprepared, head restraint and position, moving or stationary. Pre-existent health status before injury (neck pain, headache, participation problems, comorbidity, pain medication use before accident). Physical health status 2 weeks after collision (pain medication use, neck pain intensity [visual analog scale (VAS)], total cervical range of motion, high number of complaints, radicular complaints, work activities [VAS], absent from work), symptom checklist-90, diagnostic radiological or other imaging techniques (yes/no), use of soft collar.					
6. Kasch et al (2001)	Denmark emergency department	12 mo	Work capacity, as evaluated using the response to a 6-item, nonvalidated scale (1-2, 3-6)	141	21
Predictors followed: Age, gender, body mass index, pain intensity (VAS), presence/absence of nonpainful complaints (exhaustion, anxiousness, forgetfulness, sleep disturbance, irritability, impaired ability to concentrate, imbalance, dizziness, nausea, increased sensitivity to noise, tinnitus, paresthesia in upper limbs, dysphagia, blurred vision, diplopia, Millon Behavioural Health Inventory, active cervical range of motion, cervical neck muscle workload).					

(frequency counts, means and standard deviations, regression coefficients and standard error, regression coefficients and *P* values, *t* values, *U* values, or χ^2 values). The effect estimate, or statistic, of interest in this review was the odds ratio for each predictor studied.

Studies that used the duration of symptoms—most often in terms of time to insurance claim closure—as the dependent variable were excluded from this review due to the inability to pool the data with studies of ongoing symptoms at a fixed follow-up.

To avoid giving artificially greater weight to any predictors that were investigated in the same cohort but reported in multiple publications, the decision was made to score quality and extract data by cohort rather than by study, an approach similar to the one used by Scholten-Peeters et al.[29] A total of 14 studies following 11 cohorts (total n = 3193) were identified that fit our inclusion criteria (**FIGURE 1**). **TABLE 1** describes the included studies.

Quality Scoring and Data Extraction

Following the suggestion of Glass,[12] we did not exclude articles based on an arbitrary threshold of methodologic quality. Instead, we developed a scoring tool, adapted from that of Scholten-Peeters et al[29] to allow for better discrimination between levels of article quality given our specific purpose. We used the score from the tool as a moderator variable in further analyses.

The scoring tool consisted of 17 items covering the areas of patient sampling, methodology, statistical analysis, and interpretation of results (**APPENDIX A**). Two authors independently scored each paper

TABLE 1 Descriptions of Cohorts Included in the Meta-analysis (continued)

Cohort, Authors	Sampling Frame	Time to Follow-up	Outcome Collected	Sample Size	Study Quality*
7. Kivioja et al (2005)	Sweden emergency department	12 mo	"Do you have neck pain now?" (yes/no)	96	24
Predictors followed: Neck pain before accident, neck pain intensity at baseline, catastrophizing subscale of Coping Strategies Questionnaire, age, gender					
8. Nederhand et al (2003, 2004)	Netherlands emergency department	6 mo	Neck disability index (NDI)[§] score, dichotomized as no disability (<15) or disability (≥15)	84	19
Predictors followed: NDI, pain intensity (VAS), Tampa Scale of Kinesiophobia (TSK), catastrophizing (PCL-E), isometric muscle activity					
9. Radanov et al (1995a,b)	Switzerland primary care	12, 24 mo	The presence of symptoms related to the crash, unclear as to what tool used	117	17
Predictors followed: Age, gender, education, type of vocational activity, dissatisfaction with occupation, crash mechanism, patient was driver, seat belt, head restraints, seat damage, unprepared, head rotated or inclined, car stationary, patient at fault, crash assessment by patient, illness or disability worry, familiarity with symptoms of whiplash, history of head injury, history of whiplash, history of headache, time to neck pain onset, time to headache onset, initial neck pain intensity, initial headache intensity, presence of neck pain, headache, fatigability, shoulder pain, anxiety, sleep disturbance, back pain, sensitivity to noise, impaired concentration, blurred vision, irritability, dizziness, forgetfulness, swallowing difficulty, jugular pain, neck muscle tenderness, restricted range of motion, radicular irritation, radicular deficit, signs or symptoms of cranial nerve deficit, diplopia, oscillopsia, unsteadiness, vertigo, tinnitus, myelopathy, olfactory dysfunction, trigeminal-facial dysfunction, pupillomotor dysfunction, radicular deficit. Injury severity (WAD grade), multiple-symptom score, radiologic findings (misalignment of cervical curve, osteoarthrosis, signs of restricted movement). Psychosocial stress (neurotic symptoms in childhood, performance problems in school, dysfunctional family, family history of somatic illness, history of psychological or behavioral problems, current stress). Frieburg Personality inventory, score on well-being scale. Cognitive: self-rated cognitive ability, digit span, corsi block tapping, number connection test, trail-making test A and B, paced auditory serial addition test (PASAT).					
10. Soderlund et al (2000)	Sweden tertiary care	6 mo	Symptomatic or not. Nonsymptomatic fall more than half an SD below the mean pain numerical rating scale (NRS) value of the cohort	47	12
Predictors followed: Pain disability index, self-efficacy scale, coping strategies questionnaire, pain intensity VAS, cervicothoracic posture, cervical range of motion, cervicocephalic kinaesthetic sensibility.					
11. Sterner et al (2003)	Sweden emergency department, primary care	16 mo	Disability, ranked according to "none/minor" or "major," based on subjective report of interference with work, leisure activities, and activities of daily living	296	21
Predictors followed: Age, sex, education, WAD grade, accident type, frequency of prior neck complaints, prior headache, prior back complaints.					

** Highest possible score, 34.*
† Berglund et al[2] captured helplessness and locus of control at 1 month postinjury, hence these data were not included in the current analysis.
‡ Hartling et al[15] followed subjects for up to 24 months. All data were presented for 6-month follow-up only. WAD grade data were presented for 6-, 12-, 18-, and 24-month follow-ups.

to provide an index of interrater reliability of the scale using Cohen's kappa for categorical items. The kappa value of the scoring tool was 0.81 across all items. The kappa values for individual items ranged from 0.44 to 1.00, with 70% of the items reaching values of 0.75 or greater. Discrepancies in scoring were primarily due to errors in reading or interpretation of scoring criteria, and were easily settled by consensus.

Data extraction was performed by the primary author (D.W.). A structured coding scheme was constructed for each predictor pertaining to cohort identification by primary author and year, methodologic quality (low, moderate, high), sampling frame (emergency department, primary care clinic, specialist clinic, insurance claims), time to follow-up (<12 months, 12-16 months, 18-24 months), outcome captured (ongoing symptoms, ongoing disability), and geographic region (North America, United Kingdom, Scandinavia, The Netherlands, Australia). Data were entered into the Comprehensive Meta-Analysis, Version 2.0 software[4] (Biostat, Inc, Englewood, NJ) for statistical manipulation.

Effect Size Calculations

Statistical conversions were required to allow for the determination of pooled odds ratios. These conversions were complicated by the range of numerical indices used for both risk factors and outcomes across the different studies, including dichotomous categorical presentations, means and standard deviations, regression coefficients, and χ^2 values. To pool results it is necessary to synthesize findings across studies using a common effect size estimator. The procedures used to convert data, where necessary, are described in **APPENDIX B**.

Samples were assembled from various geographic regions and points of contact with the healthcare system, all statistical pooling was performed using a random-effects model, which is a more conservative approach when heterogeneity of the population is thought to exist.[16]

Missing Data

In studies that indicated the collection and analysis of a risk factor and described "no significant effect" of that risk

factor without accompanying data, we used a conservative approach and assumed equal frequency of exposure in each group of recovered/not-recovered patients. This forced the odds ratio to 1.00 (log odds ratio, 0.00) with $SE_{logoddsratio}$ (used for calculating 95% confidence intervals [CIs]) equal to the square root of the sample variance.

Moderator Analysis

It is possible that effect sizes (magnitude of the odds ratio in this review) are influenced by systematic sources of bias that can be explored separately. A moderator variable can be thought of as a stratification variable, in which data are grouped and analyzed within and between levels of the variable to determine what effect, if any, that variable has on the outcome. In this review, the presence of moderator variables was determined through evaluation of the Q statistic.[7] Specifically, we evaluated the moderating effect of 4 variables determined a priori: (1) study quality, based on our quality-scoring tool, categorized as strong (>22/34, n = 4), moderate (18-22/34, n = 4), or weak (<18/34, n = 3); (2) outcome captured, categorized as the presence of ongoing symptoms (n = 8) or ongoing disability (n = 4), with the study by Berglund et al[2] capturing both outcomes; (3) sampling frame categorized as emergency department (n = 6), primary care (n = 4), tertiary care (n = 1), or insurance claims (n = 1), with the study by Sterner et al[35] sampling from emergency departments and primary care; and (4) length of follow-up categorized as less than 12 months (n = 5), 12 to 16 months (n = 8), or 18 months or greater (n = 3). Hartling et al,[15] Berglund et al,[2] and Radanov et al[26] each collected outcomes at multiple time points, in which case, the earliest follow-up point was used. These moderators were chosen based on the possibility that they may function as confounding variables in interpretation of results. Geographic region was also considered for use as a moderator, given empirical support for this in the literature[9]; but our sample of literature was too small to perform meaningful analyses on geographic region.

The Q statistic is a statistical test of the null hypothesis that the effect sizes from each cohort in the sample are the same. The test provides a *P* value indicating the probability that the heterogeneity within the sample of effect sizes is truly greater than zero. To avoid type II error, we chose a liberal *P* value of .1 as significant for heterogeneity. For each individual predictor identified, a significant overall Q_{within} indicates substantial heterogeneity within the sample of effect sizes. In this case, the sample is categorized based on 1 of the moderator variables listed above, and the Q_{within} for each category is determined along with the $Q_{between}$ as an omnibus test of significance between the levels of the moderator variable. An appropriate moderator variable was identified when the Q_{within} for each level of the variable was nonsignificant, indicating homogeneity within levels, and the $Q_{between}$ was significant indicating heterogeneity between levels of the moderator. This procedure can be considered analogous to the *F* test in an analysis of variance.

Publication Bias

It is possible that the results of a meta-analysis are biased due to the fact that studies finding nonsignificance are less likely to be published, leading to an overestimation of the effect size in meta-analysis. To test for this bias, we examined the funnel plot for each predictor and calculated the fail-safe N statistic.[28] The fail-safe N can be considered an omnibus test of the robustness of the result, providing an estimate of the number of unpublished studies of nonsignificant results that would be required to nullify the findings of significant pooled effect size.

RESULTS

The number of predictors identified from the 11 cohorts totaled 239, averaging 22 predictors and 49 effect sizes per cohort (**TABLE 1**). A total of 535 different effect sizes were extracted, including various permutations of a single predictor (ie, WAD 2 versus WAD 1, WAD 3 versus WAD 1, WAD 3 versus WAD 2, and WAD 3 versus WAD 1 and 2) and at all time points.

The definition of chronicity varied for each cohort. Out of 11 cohorts followed, there were 13 different criteria for the presence of persistent pain and/or disability. There were 8 cohorts in which some indicator of persistent symptoms was captured, and 4 cohorts in which an indicator of persistent disability was captured, with Berglund et al[2] capturing both.

We identified 25 predictors that had been studied in at least 3 cohorts, and that had enough information presented to allow for meaningful statistical pooling. In the interest of clarity and readability, the 25 predictors have been grouped into 4 categories for presentation: patient demographics (4 predictors), collision parameters (9 predictors), previous history (2 predictors), and presenting symptoms (10 predictors).

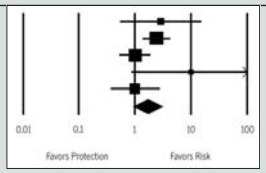
Demographic Variables

FIGURE 2 provides a graphic representation of the odds ratio and 95% CI (forest plot) for the 4 demographic variables of older age (n = 1142), no postsecondary education (n = 2019), female gender (n = 3109), and obese body mass index (BMI) category (n = 559).

Four studies[17,18,30,35] indicated older age as a variable, but did not provide a clear definition of "older." Three studies[3,15,26] indicated that older age referred to subjects over the age of 50, compared with subjects under 50. There was significant heterogeneity in the pooling of all studies (Q = 16.7, *P* = .025) for older age, so a moderator analysis was performed. When the studies were stratified by type of outcome (persistent symptoms versus disability), the Q_{within} became nonsignificant while the $Q_{between}$ remained significant, indicating an appropriate moderator was found. As seen in **FIGURE 2**, the size of the effect of older age on the risk of persistent disability is negligible (odds ratio [OR], 0.99; 95% CI: 0.97-1.01). The size of the effect of older age on persistent pain is stronger

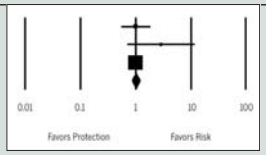
A

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Borchgrevink (1997)	Pain	6	2.85	0.54	15.00	1.24	0.22	
Hartling (2002)	Pain	6	2.40	1.37	4.22	3.04	0.00	
Hendriks (2005)	Pain	12	1.00	0.53	1.90	0.00	1.00	
Radanov (1995)	Pain	24	10.00	0.86	115.94	1.84	0.07	
Soderlund (2000)	Pain	6	1.00	0.37	2.73	0.00	1.00	
			1.68	0.93	3.05	1.72	0.09	



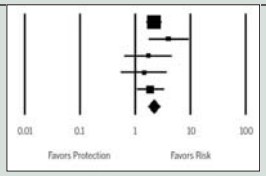
B

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Kasch (2001)	Disability	12	1.00	0.55	1.83	0.00	1.00	
Nederhand (2004)	Disability	6	2.90	0.71	11.83	1.48	0.14	
Sterner (2003)	Disability	12	0.99	0.97	1.01	−0.97	0.33	
			1.01	0.84	1.21	0.09	0.93	



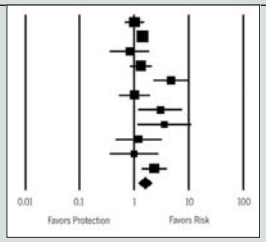
C

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Berglund (2006)	Pain	6	2.19	1.59	3.00	4.84	0.00	
Hendriks (2005)	Pain	12	4.01	1.75	9.24	3.27	0.00	
Kivioja (2005)	Pain	12	1.71	0.63	4.65	1.06	0.29	
Radanov (1995)	Pain	24	1.43	0.54	3.77	0.73	0.47	
Sterner (2003)	Disability	12	1.87	1.05	3.32	2.14	0.03	
			2.15	1.68	2.75	6.09	0.00	



D

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	1.00	0.66	1.50	−0.01	0.99	
Berglund (2006)	Pain	6	1.40	1.11	1.76	2.82	0.00	
Borchgrevink (1997)	Pain	6	0.82	0.35	1.89	−0.47	0.64	
Hartling (2002)	Pain	6	1.31	0.82	2.12	1.12	0.26	
Hendriks (2005)	Pain	12	4.73	2.23	10.00	4.06	0.00	
Kasch (2001)	Disability	12	1.00	0.52	1.93	0.00	1.00	
Kivioja (2005)	Pain	12	2.99	1.18	7.55	2.31	0.02	
Nederhand (2004)	Disability	6	3.52	1.13	10.95	2.17	0.03	
Radanov (1995)	Pain	24	1.21	0.46	3.19	0.39	0.70	
Soderlund (2000)	Pain	6	1.00	0.36	2.81	0.00	1.00	
Sterner (2003)	Disability	16	2.32	1.37	3.93	3.12	0.00	
			1.54	1.16	2.06	2.96	0.00	



E

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Hartling (2002)	Pain	6	0.96	0.59	1.55	−0.18	0.85	
Kasch (2001)	Disability	12	1.00	0.55	1.83	0.00	1.00	
Nederhand (2004)	Disability	6	3.00	1.13	7.98	2.20	0.03	
			1.24	0.71	2.19	0.76	0.45	

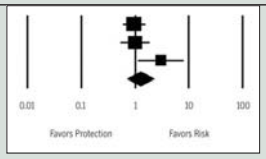
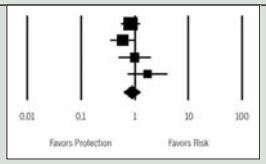


FIGURE 2. Demographic variables that have been investigated at least 3 times as risk factors for persistent whiplash-related problems. (A) older age as a risk factor for persistent pain. (B) Older age as a risk factor for persistent disability. (C) No postsecondary education. (D) Female gender. (E) Body Mass Index (obese). For all variables and figures, the square represents the odds ratio for that study and the lines are the 95% confidence interval. The diamond represents the pooled odds ratio from all studies, with the upper and lower bounds of the 95% confidence interval represented by the right and left corners of the diamond, respectively. If the confidence interval crosses 1, the variable is not a significant predictor of outcome.

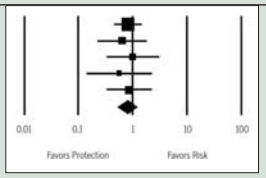
A

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	0.84	0.56	1.26	−0.84	0.40	
Hartling (2002)	Pain	6	0.60	0.36	1.02	−1.88	0.06	
Hendriks (2005)	Pain	12	1.00	0.50	2.02	0.00	1.00	
Radanov (1995)	Pain	12	1.73	0.73	4.11	1.24	0.21	
			0.87	0.61	1.25	−0.75	0.45	



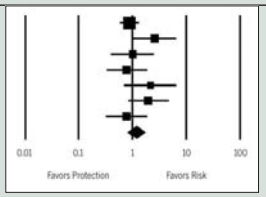
B

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	0.81	0.46	1.44	−0.71	0.48	
Borchgrevink (1997)	Pain	6	0.64	0.23	1.82	−0.84	0.40	
Hendriks (2005)	Pain	12	1.00	0.33	3.04	0.00	1.00	
Nederhand (2004)	Disability	6	0.56	0.14	2.24	−0.82	0.41	
Radanov (1995)	Pain	12	0.85	0.32	2.26	−0.32	0.75	
			0.79	0.53	1.17	−1.18	0.24	



C

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	0.86	0.57	1.29	−0.73	0.46	
Borchgrevink (1997)	Pain	6	2.56	1.03	6.39	2.02	0.04	
Hendriks (2005)	Pain	12	1.00	0.40	2.51	0.00	1.00	
Kivioja (2005)	Pain	12	0.77	0.32	1.84	−0.59	0.56	
Nederhand (2004)	Disability	6	2.14	0.70	6.58	1.33	0.18	
Radanov (1995)	Pain	12	1.97	0.83	4.65	1.54	0.12	
Sterner (2003)	Disability	12	0.77	0.31	1.89	−0.57	0.57	
			1.17	0.81	1.70	0.84	0.40	



D

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	1.63	1.01	2.62	2.00	0.05	
Borchgrevink (1997)	Pain	6	0.14	0.03	0.67	−2.45	0.01	
Hendriks (2005)	Pain	12	1.00	0.33	3.04	0.00	1.00	
Nederhand (2004)	Disability	6	0.47	0.09	2.38	−0.91	0.36	
Radanov (1995)	Pain	12	0.45	0.10	2.13	−1.01	0.31	
			0.66	0.27	1.59	−0.93	0.35	

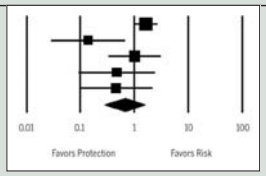


FIGURE 3. Collision parameters and their influence on the risk of persistent whiplash-related problems at follow-up. (A) Vehicle stationary when hit. (B) Frontal collision. (C) Rear-end collision. (D) Side-on or "other" collision. (E) Driver of vehicle (compared to other positions in vehicle). (F) Front passenger in vehicle. (G) Unprepared for the collision. (H) No seat belt in use at time of collision. (I) No head rest in use at time of collision.

(OR, 1.68; 95% CI: 0.93-3.06), but failed to reach statistical significance.

Obesity, as indicated by a BMI greater than 27, was investigated in 3 cohorts.[18,5,22] The results were homogenous and indicated no significant effect of BMI on outcome (OR, 1.24; 95% CI: 0.71-2.19).

Lower education, defined here as having no postsecondary education, was investigated in 5 cohorts[2,17,19,26,35] and appears to be a risk factor for persistent WAD-related pain or disability (OR, 2.15; 95% CI: 1.69-2.75), with homogeneity.

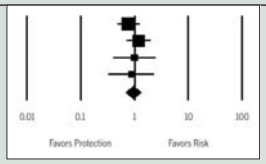
Female gender as a risk factor, investigated in all 11 cohorts, showed a significant but modest effect (OR, 1.54; 95% CI: 1.16-2.06). There was significant heterogeneity within the pool of effect sizes (Q, 24.3; $P<.01$). We were unable to identify a logical or meaningful moderator variable.

Collision Parameters

FIGURE 3 provides the forest plots for the 9 collision parameter variables: vehicle stationary when hit (n = 1050), frontal collision (n = 888), rear-end collision (n = 984), side or "other" collision (n = 800), driver of vehicle (n = 1050), front passenger in vehicle (n = 931), unprepared for the collision (n = 1050), no seat belt used (n = 1050), and no head restraint used (n = 1050). The only variable that proved to have significant predictive power was that of not wearing a seat belt at the time of collision, collected in 4 cohorts.[1,15,17,26] Not wearing a seat belt led to a nearly 2-fold

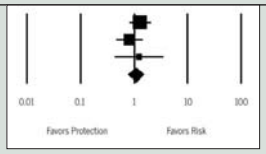
E

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Atherton (2006)	Pain	12	0.77	0.49	1.23	−1.08	0.28
Hartling (2002)	Pain	6	1.21	0.72	2.03	0.71	0.48
Hendriks (2005)	Pain	12	1.00	0.40	2.51	0.00	1.00
Radanov (1995)	Pain	12	0.87	0.32	2.33	−0.29	0.77
			0.94	0.69	1.28	−0.38	0.71



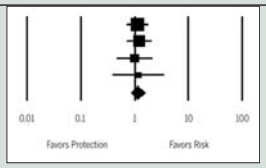
F

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Atherton (2006)	Pain	12	1.29	0.81	2.06	1.08	0.28
Hartling (2002)	Pain	6	0.83	0.47	1.44	−0.68	0.50
Radanov (1995)	Pain	12	1.24	0.43	3.56	0.41	0.68
			1.09	0.78	1.53	0.50	0.62



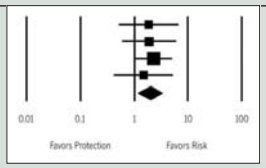
G

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Atherton (2006)	Pain	12	1.14	0.72	1.80	0.56	0.58
Hartling (2002)	Pain	6	1.21	0.70	2.09	0.68	0.49
Hendriks (2005)	Pain	12	1.00	0.47	2.12	0.00	1.00
Radanov (1995)	Pain	24	1.19	0.40	3.57	0.31	0.76
			1.14	0.84	1.55	0.84	0.40



H

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Atherton (2006)	Pain	12	1.86	0.52	6.70	0.95	0.34
Hartling (2002)	Pain	6	1.88	0.59	5.95	1.07	0.29
Hendriks (2005)	Pain	12	2.28	1.05	4.94	2.09	0.04
Radanov (1995)	Pain	12	1.48	0.42	5.24	0.61	0.54
			1.97	1.17	3.32	2.54	0.01



I

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Atherton (2006)	Pain	12	0.99	0.38	2.56	−0.02	0.98
Hartling (2002)	Pain	6	0.84	0.52	1.38	−0.68	0.50
Hendriks (2005)	Pain	12	1.00	0.47	2.12	0.00	1.00
Radanov (1995)	Pain	24	1.16	0.23	5.89	0.18	0.86
			0.91	0.63	1.32	−0.48	0.63

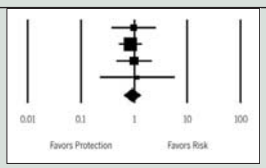


FIGURE 3. *Continued*

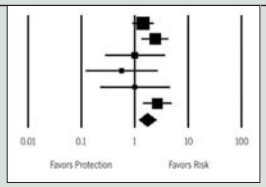
increase in the risk of developing persistent WAD-related pain or disability at follow-up (OR, 1.97; 95% CI: 1.17-3.32).

Past Medical History

FIGURE 4 provides the forest plots for the 2 variables pertaining to past medical history: preaccident history of neck pain (n = 1393) and preaccident history of headache (n = 532). It should be noted that the criteria for determining whether a patient did or did not have a past history of neck pain or headache was not clear, and were collected through patient self-report, raising the possibility of recall bias. The effect size for history of neck pain was homogenous, and indicated a small but significant risk of developing persistent WAD-related problems at follow-up (OR, 1.70; 95% CI: 1.17-2.48). The effect size for history of headache demonstrated significant heterogeneity (Q = 8.9, P = .012), with the effect from Radanov et al[26] the clear outlier. With data from only 3 cohorts it was difficult to identify a moderator variable, but the difference in length of follow-up between Radanov et al[26] (24 months), Hendriks et al[17] (12 months), and Sterner et al[35] (16 months) might explain the difference. Regardless, the effect of previous

A

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	1.41	0.88	2.25	1.42	0.15	
Hartling (2002)	Pain	6	2.39	1.34	4.25	2.95	0.00	
Hendriks (2005)	Pain	12	1.00	0.27	3.64	0.00	1.00	
Radanov (1995)	Pain	24	0.57	0.12	2.70	−0.71	0.48	
Soderlund (2000)	Pain	6	1.00	0.22	4.49	0.00	1.00	
Sterner (2003)	Disability	16	2.63	1.42	4.88	3.07	0.00	
			1.70	1.17	2.48	2.75	0.01	



B

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Hendriks (2005)	Pain	12	1.00	0.37	2.71	0.00	1.00	
Radanov (1995)	Pain	24	10.00	2.75	36.34	3.50	0.00	
Sterner (2003)	Disability	16	1.35	0.74	2.46	0.98	0.33	
			2.15	0.69	6.73	1.31	0.19	

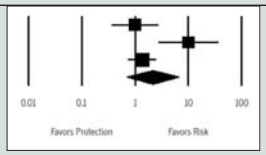


FIGURE 4. The influence of preaccident history of (A) neck pain and (B) headache on the risk of persistent whiplash-related problems at follow-up.

history of headache on the risk of ongoing problems at follow-up, calculated from all 3 cohorts, failed to reach significance (OR, 2.15; 95% CI: 0.69-6.73).

Presenting Signs and Symptoms

FIGURE 5 provides the forest plots for the 8 variables pertaining to presenting signs and symptoms: high neck pain intensity (n = 2391), restricted cervical range of motion (n = 904), report of disturbed sleep since the accident (n = 570), current presence of headache (n = 1930), current presence of neck pain (n = 539), current presence of radicular symptoms (n = 570), higher catastrophic cognitions (n = 277), and depressive symptoms (n = 541). The strongest predictive variables were those that dealt with pain on presentation. The presence of headache (**FIGURE 5D**) or neck pain (**FIGURE 5E**) at initial assessment was associated with a significant increase in the risk of reporting persistent WAD-related problems at follow-up (headache OR, 2.71; 95% CI: 2.16-3.41; neck pain OR, 2.87; 95% CI: 1.51-5.46).

Data on the intensity of neck pain at intake, captured in 8 cohorts,[2,15,17-19,22,26,30] were presented in such a way that it was possible to establish a meaningful and consistent cut point of 55 out of 100, or 5 out of 10, on a pain visual analog scale or numeric rating scale, respectively, for each cohort. Those subjects who reported a pain intensity at intake of greater than 55/100 demonstrated a nearly 6-fold increase in the risk of persistent pain or disability at follow-up (OR, 5.77; 95% CI: 2.89-11.52). Moderator analysis of these data revealed significant heterogeneity of effect sizes (Q = 28.4; $P<.001$). Using type of outcome (pain versus disability) as a moderator variable, the odds ratios from the cohorts that used disability as an outcome became homogenous, while significant heterogeneity remained amongst the pooled odds ratios when symptoms were the outcome. The odds ratio, when ongoing symptoms were used as the outcome, was 5.14 (95% CI: 2.14-12.38) and 12.57 (95% CI: 8.66-18.26) when ongoing disability was the outcome. This difference failed to reach statistical significance ($Q_{between}$ = 2.20; $P>.10$), and therefore type of outcome (pain versus disability) did not fully meet the criteria to be deemed an adequate moderator variable[7] in this analysis.

Catastrophic cognitions were captured in 3 cohorts.[19,22,30] Catastrophizing in 2 cohorts[19,30] was captured using the catastrophizing subscale of the coping strategies questionnaire.[27] We were able to establish a meaningful cut point of greater than 7 points out of the total 36 points possible on the scale. Nederhand et al[22] used the Pain Cognition List-Experimental (PCL-E)[39] to capture catastrophizing. These authors used the area under the receiver operating characteristic curve to identify a meaningful cut point of 15 points on the catastrophizing subscale of the tool. Perhaps due to the use of 2 different, albeit valid, measures of catastrophizing, the effect sizes were heterogeneous (Q = 5.8; P = .055). The findings indicated that high catastrophizing is a significant risk for poor outcome (OR, 3.77; 95% CI: 1.33-10.74). No clear moderator could be found with only 3 cohorts in the sample.

The presence of depressive symptoms was based on scores on the SCL-90,[17] the MCMI-1,[3] or through telephone interview.[15] We were able to perform a meaningful statistical pooling procedure here based solely on the fact that there was no significant effect for depression when the SCL-90 or MCMI-1 were used, so the effect size was forced to nil. The dichotomization procedure would otherwise not have been possible, as insufficient data were presented. This indicates that, based on limited evidence, depressive symptoms appear to play no role in outcome following WAD.

TABLE 2 Summary of Key Findings

	Relative Increase in Risk Over Someone Without the Factor (95% CI)	Number of Independent Cohorts (Total n Subjects)	Fail-Safe N*	Mean Study Quality†
Variables with strong evidence of a significant effect‡				
High baseline neck pain intensity (greater than 55/100)	5.72 (2.95, 11.10)	8 (2391)	310	21.0
Presence of headache on intake	2.71 (2.16, 3.41)	4 (1930)	53	20.0
WAD grade 3 (compared to grade 2, when outcome is captured 12 mo postinjury)	2.41 (1.62, 3.59)	3 (2205)	16	24.0
No postsecondary education	2.15 (1.69, 2.75)	5 (2019)	33	23.0
WAD grade 2 or 3 (compared to grade 0 or 1, when outcome is captured 12 mo postinjury)	1.96 (1.41, 2.74)	4 (2501)	29	23.5
Variables with moderate evidence of a significant effect§				
Catastrophizing (>7 on CSQ subscale or >15 on PCL-E)	3.77 (1.33, 10.74)	3 (277)	11	18.0
Presence (yes/no) of neck pain at intake	2.87 (1.51, 5.46)	3 (539)	5	17.5
No seat belt in use at time of accident	1.97 (1.17, 3.32)	4 (1050)	2	22.5
History of neck pain prior to the accident	1.70 (1.17, 2.48)	6 (1393)	6	20.5
Female	1.54 (1.16, 2.06)	11 (3109)	25	21.0
Variables that just miss significance‖				
Disturbed sleep	2.96 (0.97, 9.04)	3 (570)		21.5
Older age (for pain as an outcome)	1.68 (0.93, 3.06)	5 (1142)		18.5
Variables with strong evidence of no effect¶				
Rear-end collision	1.17 (0.81, 1.70)	7 (984)		21.0
Unprepared-for collision	1.14 (0.84, 1.55)	4 (1050)		22.5
Front passenger in vehicle	1.09 (0.78, 1.53)	3 (931)		21.0
Driver of vehicle	0.94 (0.69, 1.28)	4 (1050)		22.5
No head rest in use	0.92 (0.63, 1.32)	4 (1050)		22.5
Vehicle stationary when hit	0.87 (0.61, 1.25)	4 (1050)		22.5
Frontal collision	0.79 (0.53, 1.17)	5 (888)		20.5
Side-on or "other" collision	0.66 (0.27, 1.59)	5 (800)		20.5

Abbreviations: CSQ, Coping Strategies Questionnaire; PCL-E, pain cognitions list, experimental; WAD, whiplash-associated disorder.
** The number of studies showing no significant effect of the predictor that would have to be included in this analysis to nullify these findings of significance. Fail-safe N cannot be calculated for factors that are not significant.*
† The highest possible score is 34.
‡ Estimate of fail-safe N at least 5 times greater than the number of studies included in the review.
§ Estimate of fail-safe N less than 5 times greater than the number of studies included in the review.
‖ Lower limit of 95% CI between 0.9 and 1.0.
¶ Investigated in at least 4 studies, upper and lower limits of 95% CI >0.10 away from 1.0.

WAD Grades

Given the presentation of the data, we were able to perform 2 meaningful comparisons using the WAD grades. A total of 6 studies evaluated WAD grade as a variable.[1,2,15,26,30,35] Hartling et al[15] and Berglund et al[2] each captured outcome at 3 different time points, which allowed for an estimation of the effect size at each time point. **FIGURE 6** shows the forest plots for the comparison of WAD grades 2 and 3 against grades 0 and 1. The size of the effect was significant and relatively consistent at each of 6 months (OR, 2.40; 95% CI: 1.41-4.10; n = 1772), 12-16 months (OR, 1.96; 95% CI: 1.41-2.74; n = 2501), and 24 months (OR, 2.67; 95% CI: 1.99-3.58; n = 1842). There was homogeneity in the effect sizes within each time point, except for the 6-month follow-up (Q = 5.0; P = .084), for which no clear moderator was identified.

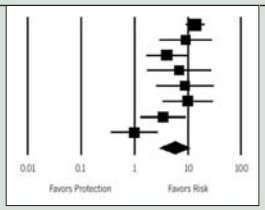
FIGURE 7 shows the forest plots when WAD grade 3 was taken as the risk factor against WAD grade 2 as the comparator. Five studies presented data that allowed for this comparison (total n = 2369). There was homogeneity of effect sizes within each time point. Here the effect decreases slightly with greater time from intake to follow-up, and becomes nonsignificant by 24-month follow-up (6 months OR, 2.66; 95% CI: 1.54-4.58; 12 months OR, 2.41; 95% CI: 1.62-3.59; 24 months OR, 1.63; 95% CI: 0.41-6.51). Of interest, the work of Radanov et al,[26] who found WAD grade 3 to have a protective effect when compared against WAD grade 2, was the only study to report such a finding.

Publication Bias

TABLE 2 provides the fail-safe N values for

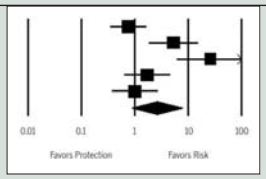
A

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Berglund (2006)	Disability	6	13.73	9.07	20.79	12.37	0.00	
Hartling (2002)	Pain	6	9.14	2.92	28.61	3.80	0.00	
Hendriks (2005)	Pain	12	4.06	1.69	9.74	3.13	0.00	
Kasch (2001)	Disability	12	6.86	1.71	27.46	2.72	0.01	
Kivioja (2005)	Pain	12	8.84	2.54	30.72	3.43	0.00	
Nederhand (2004)	Disability	6	9.99	3.38	29.49	4.17	0.00	
Radanov (1995)	Pain	24	3.41	1.29	9.01	2.47	0.01	
Soderlund (2000)	Pain	6	1.00	0.37	2.73	0.00	1.00	
			5.72	2.95	11.10	5.15	0.00	



B

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	0.74	0.34	1.61	−0.75	0.45	
Hendriks (2005)	Pain	12	5.29	1.83	15.26	3.08	0.00	
Kasch (2001)	Disability	12	26.22	6.13	112.21	4.40	0.00	
Radanov (1995)	Pain	24	1.69	0.63	4.56	1.04	0.30	
Soderlund (2000)	Pain	6	1.00	0.37	2.73	0.00	1.00	
			2.56	0.85	7.72	1.66	0.10	



C

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Hartling (2002)	Pain	6	4.22	2.56	6.94	5.66	0.00	
Hendriks (2005)	Pain	12	1.00	0.53	1.90	0.00	1.00	
Radanov (1995)	Pain	24	7.04	2.36	21.00	3.50	0.00	
			2.96	0.97	9.04	1.90	0.06	

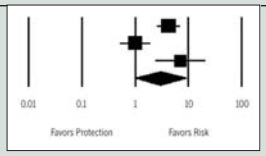


FIGURE 5. Variables pertaining to presenting signs and symptoms and their influence on the risk of persistent whiplash-related problems at follow-up. (A) Neck pain intensity at baseline >55/100. (B) Restricted cervical range of motion. (C) Disturbed sleep. (D) Presence of headache at intake. (E) Presence of neck pain at intake. (F) Presence of radicular symptoms at intake. (G) Presence of depressive symptoms. (H) Catastrophizing (see text for details).

each of the significant predictors identified above. Rosenthal[28] proposes that the fail-safe N should be at least 5 times greater than the number of studies in the analysis before an effect can be considered robust to publication bias. Based on that criterion, no postsecondary education, high neck pain intensity, report of headache at intake, and WAD grades can be considered findings that are robust to publication bias.

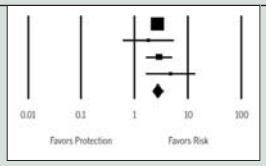
DISCUSSION

Using a methodical, stepwise approach to search the literature, perform quality assessment, and extract data, we have identified a relatively homogenous subset of cohorts from the prognostic WAD literature that allowed for meaningful statistical pooling. Based on the results of our meta-analysis, there are a number of factors for which information is easy to obtain clinically that have the potential to provide the clinician with an overall estimate of a patient's level of risk for persistent WAD-related problems at least 6 months following the injury (**TABLE 2**).

In agreement with the findings of 2 previous narrative reviews on the topic,[8,29] we have found that a rating of high neck pain intensity at intake, here defined as greater than 55/100 on a visual analog scale, is the strongest predictor of ongoing problems at long-term follow-up. Our study builds on this work by indicating the extent of this risk to be nearly a 6-fold increase in the level of risk of ongoing pain or disability. In contrast to the 2 previous narrative reviews, we were able to identify 8 additional significant risk factors for persistent problems using a statistical-pooling technique. Other significant risk factors from the currently available literature are broadly classed as demographic variables (no postsecondary education, female gender) and variables pertaining to initial signs and symptoms (presence of headache or neck pain, previous history of neck pain, catastrophizing, WAD grade). The only collision-related variable to show any predictive value when captured within 3 weeks of the collision is the nonuse of a seat belt, but it is a weak finding (OR, 1.97; fail-safe N = 2). These findings appear to support a biopsychosocial model for the development of persistent pain, but it should be noted that only 5 studies[17,18,22,25,30] included a

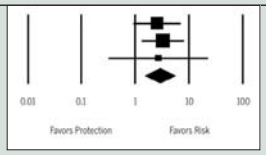
D

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Berglund (2006)	Pain	12	2.65	2.03	3.45	7.17	0.00
Borchgrevink (1997)	Pain	6	1.78	0.59	5.34	1.03	0.30
Hartling (2002)	Pain	6	2.89	1.64	5.08	3.68	0.00
Radanov (1995)	Pain	12	4.70	1.64	13.48	2.88	0.00
			2.71	2.16	3.41	8.52	0.00



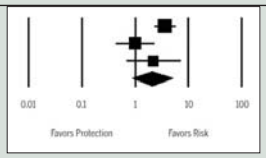
E

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Barchgrevink (1997)	Pain	6	2.51	0.91	6.90	1.78	0.08
Hartling (2002)	Pain	6	3.25	1.32	8.01	2.56	0.01
Radanov (1995)	Pain	12	2.67	0.32	22.31	0.91	0.37
			2.87	1.51	5.46	3.22	0.00



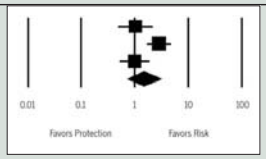
F

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Hartling (2002)	Pain	6	3.64	2.27	5.82	5.38	0.00
Hendriks (2005)	Pain	12	1.00	0.43	2.32	0.00	1.00
Radanov (1995)	Pain	24	2.19	0.68	7.06	1.31	0.19
			2.09	0.88	4.95	1.68	0.09



G

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Borchgrevink (1997)	Pain	6	1.02	0.48	2.18	0.05	0.96
Hartling (2002)	Pain	6	2.79	1.67	4.65	3.92	0.00
Hendriks (2005)	Pain	12	1.00	0.53	1.90	0.00	1.00
			1.47	0.71	3.03	1.03	0.30



H

Study Name	Outcome	Follow-up (mo)	Odds Ratio	Lower Limit	Upper Limit	z Value	P Value
			Statistics for Each Study				
Kivioja (2005)	Pain	12	1.81	0.75	4.35	1.33	0.18
Nederhand (2004)	Disability	6	9.99	3.38	29.49	4.17	0.00
Soderlund (2000)	Pain	6	3.23	0.91	11.42	1.82	0.07
			3.77	1.33	10.74	2.49	0.01

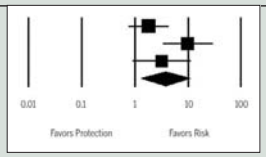


FIGURE 5. *Continued*

physical assessment in their battery of potential predictors, and these were not standardized across studies. This may account for the lack of support for physical findings in this analysis.

Four variables demonstrate robustness to publication bias through evaluation of the fail-safe N statistic: no postsecondary education, high neck pain intensity, the presence of headache, and WAD grade 2 or 3.

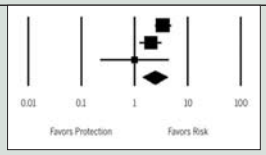
Self-rated anxiety at baseline was captured in 4 cohorts,[3,15,17,26] but the data were not presented in such a way as to allow clinically or statistically meaningful comparison. It is unfortunate that a more standardized approach to capturing post-accident anxiety has not been adopted, as there is good theoretical evidence that anxiety and/or fear play a key role in the development of persistent pain or disability.[40]

We identified several variables that were investigated in 2 cohorts, and even more that were investigated in only 1 cohort. It is our belief that statistical pooling with data from only 2 cohorts lacks meaning and provides potentially misleading results, especially where heterogeneity of effect sizes exists, so we have excluded those variables from this discussion. Many of the cognitive and perceptual variables, such as coping and threat appraisal, were among these.

It should be recognized that systematic review and meta-analysis are susceptible to publication bias, insofar as data for pos-

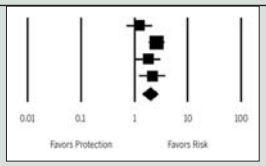
A

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Berglund (2006)	Pain	6	3.42	2.39	4.89	6.73	0.00	
Hartling (2002)	Pain	6	2.00	1.24	3.20	2.87	0.00	
Soderlund (2000)	Pain	6	1.00	0.23	4.43	0.00	1.00	
			2.40	1.41	4.10	3.22	0.00	



B

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	1.23	0.71	2.13	0.73	0.47	
Berglund (2006)	Pain	12	2.61	1.88	3.62	5.71	0.00	
Hartling (2002)	Pain	12	1.80	1.04	3.10	2.11	0.03	
Sterner (2003)	Disability	16	2.17	1.23	3.83	2.67	0.01	
			1.96	1.41	2.74	3.97	0.00	



C

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Berglund (2006)	Pain	24	2.72	1.97	3.77	6.04	0.00	
Hartling (2002)	Pain	24	2.63	1.23	5.63	2.49	0.01	
Radanov (1995)	Pain	24	1.76	0.37	8.35	0.71	0.48	
			2.67	1.99	3.58	6.54	0.00	

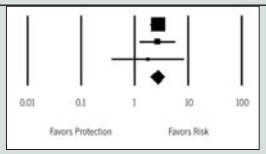


FIGURE 6. The influence of whiplash-associated disorders grade 2 or 3 (as risk factor) against grades 0 or 1 (as comparator) on the risk of persistent whiplash-related problems, at follow-ups of (A) 6 months, (B) 12 to 16 months, and (C) 24 months.

itive results are more likely to be published than are negative results. The calculation of the fail-safe N for significant findings lends confidence to the robustness of the findings from this review. However, it is possible that there are other factors that have been followed in 3 cohorts but have been excluded from the meta-analysis because the results, or data adequate for pooling, were not published. Readers should be aware of this limitation.

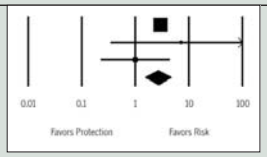
Medico-legal factors were not well represented in this sample. The criteria for inclusion of articles with inception times of 3 weeks or less probably means that many subjects would not have entered into formal litigation when baseline predictors were collected. Only Hendriks et al[17] investigated the effect of retaining a lawyer or having private insurance. Neither factor (retaining a lawyer nor having insurance) demonstrated a significant ability to predict outcome. It has been suggested that the medico-legal setting plays a significant role in recovery from WAD,[9] and this item should be considered for inclusion in future studies of this type.

In an effort to identify a homogenous pool of literature, we included only those studies in which all subjects in the cohort were intercepted within 3 weeks of the accident. We chose the 3 week cut-off as an acceptable definition of acute pain[21] and because it is within the first 3 weeks following injury, during which physical therapists are often asked to comment on the prognosis of the patient.[10] We considered the potential impact of this decision in terms of excluding potentially useful additional data. The clear strength is in the ability to draw clinically meaningful conclusions from a more homogenous pool of literature. Clinicians are able to use the results presented here in a meaningful way by applying them to only the patient population seen within the first 3 weeks of injury. The weakness is that we exclude potentially useful cohorts in which not all subjects were recruited within 3 weeks of injury. The most significant cohort that we have excluded in this case is that of Sterling et al,[32,34] who evaluated 78 subjects within 4 weeks of injury to determine the predictive ability of a multitude of physical and psychological variables. While this cohort just missed our inclusion criteria, the findings from that cohort are largely consistent with the findings of this review. Had this cohort been included in the current review, it would have impacted the effect of older age only. A sensitivity analysis, in which we included the data from the Sterling et al[32,34] cohort, showed a slight increase in effect size of older age on persistent disability that was neither statistically nor clinically significant (OR, 1.06; 95% CI: 0.93-10.20).

Some of the factors used in the analysis were presented as continuous data. This required the identification of a meaningful cut point and a dichotomization of these data for the purpose of meaningful pooling (**APPENDIX B**). The process of dichotomization of continuous data, as done in this analysis, assumes a normal distribution of data. We

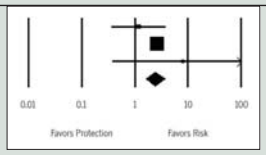
A

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Berglund (2006)	Pain	6	2.94	2.24	3.85	7.81	0.00	
Hartling (2002)	Pain	6	7.14	0.34	150.71	1.26	0.21	
Soderlund (2000)	Pain	6	1.00	0.23	4.43	0.00	1.00	
			2.66	1.54	4.58	3.52	0.00	



B

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Atherton (2006)	Pain	12	1.18	0.37	3.80	0.28	0.78	
Berglund (2006)	Pain	12	2.57	1.96	3.38	6.77	0.00	
Hartling (2002)	Pain	12	7.97	0.37	169.42	1.33	0.18	
			2.41	1.62	3.59	4.34	0.00	



C

Study Name	Outcome	Follow-up (mo)	Statistics for Each Study: Odds Ratio	Lower Limit	Upper Limit	z Value	P Value	Odds Ratio and 95% CI
Berglund (2006)	Pain	24	2.74	2.09	3.59	7.28	0.00	
Hartling (2002)	Pain	24	3.34	0.13	85.56	0.73	0.47	
Radanov (1995)	Pain	24	0.32	0.04	2.62	−1.06	0.29	
			1.63	0.41	6.51	0.69	0.49	

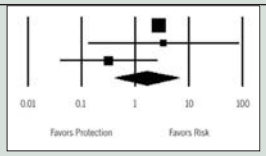


FIGURE 7. The influence of whiplash-associated disorders grade 3 (as risk factor) against grade 2 (as comparator) on the risk of persistent whiplash-related problems, at follow-ups of (A) 6 months, (B) 12 months, and (C) 24 months.

have taken every reasonable step to ensure that the data presented in the included studies were not skewed by outliers; but it remains a possibility that the dichotomization procedure overestimates or underestimates the true odds ratio. By scouring each report and using 2 independent methods of statistical conversion to come to the same result (**APPENDIX B**), we are confident that our findings represent a reasonable estimation of the true odds ratio as presented in the literature for those predictors in which the dichotomization procedure was used.

Despite some controversy in the literature,[11] it appears that the WAD classification (1 through 4) system provides a reasonable estimate of risk. Unfortunately, relying solely on the WAD system does not provide any information as to the nature or mechanism behind that risk. Several authors have proposed modifications to the WAD classification system[14,32,37] based on findings from their own cohorts, or have proposed new risk assessment tools altogether.[13,15,25] None of these tools or recommendations have, to date, been tested on cohorts independent from the one on which they were developed. Our findings suggest that simple modifications to the WAD classification system are warranted. Additional cohort studies that evaluate these changes as well as other classification systems would be useful to help clinicians and could be used to develop more specific clinical pathways or clinical practice guidelines.

There is a clear need for a standardized definition of "chronic" WAD. From the 11 cohorts identified for this review, there were 13 different criteria for identifying the presence of persistent problems. We chose only those cohorts in which the outcome was expressed as ongoing pain or ongoing disability. Of interest is that for none of the significant predictors did type of outcome (ongoing pain or ongoing disability) significantly affect the size of the odds ratio, although there appears to be a trend for type of outcome to moderate the effect of pain intensity. This might suggest that all of the outcomes, while seemingly different, captured a similar underlying construct.

Odds ratios from observational studies can be used to provide an estimate of the relative increase in risk of a poor outcome for a patient with a specific risk factor, as compared to another individual who does not have that factor, but the absolute values should be interpreted with caution. Further, it is unclear as to how these odds ratios can be combined for the patient with several risk factors beyond an appreciation that some factors represent greater risk than others, and the more risk factors present the greater the risk of poor long-term outcome. It is important for clinicians to become familiar with the factors that increase risk of poor outcomes and the extent of increased risk to provide more accurate prognosis to their patients. Some risk factors will be modifiable and others will not. Risk modification studies or studies that evaluate the response to different interventions on the basis of these predictors will provide evidence that allows clinicians to not only predict

outcomes more accurately but to customize treatment plans.

There are 2 key areas that demand clinicians exercise caution in interpreting the results of this analysis. First, it must be noted that the overall quality of the literature from which the data were extracted was moderate, with some common threats to internal validity, such as a lack of clear validity for the method of capturing many of the prognostic variables, or the lack of blinded assessors. Second, the bivariate nature of the relationships between exposure and outcome does not control for the potential interaction between variables. For example, it might be the case that people with no postsecondary education tend to catastrophize about pain more than those with higher levels of education, thereby making education level significant only by virtue of its association with catastrophizing. The nature of the data did not allow for metaregression, and thus the interaction between predictor variables cannot be evaluated.

In summary, based on the results of our meta-analysis of the available prognostic literature in WAD, we make the following recommendations. (1) The initial assessment should document the presence of the factors listed in the top 2 sections of **TABLE 2**, and note the extent to which each increase the risk of an adverse outcome at 6 to 24 months. (2) We recommend that documentation, particularly documentation provided to payers, reference this meta-analysis to indicate the source of evidence and provide an indication of evidence-based practice. (3) We recommend that the following statement be included wherever statements listed in 1 are used: "Patients often have 1 or more risk factors that affect their outcome. The evidence available at this time indicates how to interpret single risk factors in the way we have in this assessment. The evidence does not provide direction on how to interpret multiple factors at this time. It cannot be assumed that multiple risk factors are directly additive, but evidence suggests that having multiple risk factors does increase risk in most cases."

CONCLUSIONS

A meta-analysis of cohort studies examining risk of persistent (at least 6 months postinjury) pain or disability in patients assessed within 3 weeks of whiplash injury indicates that several risk factors should be routinely assessed during history taking and considered when establishing a prognosis for recovery. Clinicians should also remember that "lack of evidence" is not synonymous with "evidence against," a point that is especially relevant when considering the absence of hard physical signs as significant predictors in our review. ◉

ACKNOWLEDGEMENTS: *The authors would like to thank Dr Charlie Goldsmith and Mr Larry Stitt for their statistical support, and Jennifer Otten, Jennifer Nelson, Melinda Genereaux, and Kimberly Taylor for their help in the literature search.*

REFERENCES

1. Atherton K, Wiles NJ, Lecky FE, et al. Predictors of persistent neck pain after whiplash injury. *Emerg Med J.* 2006;23:195-201. http://dx.doi.org/10.1136/emj.2005.027102
2. Berglund A, Bodin L, Jensen I, Wiklund A, Alfredsson L. The influence of prognostic factors on neck pain intensity, disability, anxiety and depression over a 2-year period in subjects with acute whiplash injury. *Pain*. 2006;125:244-256. http://dx.doi.org/10.1016/j.pain.2006.05.026
3. Borchgrevink GE, Stiles TC, Borchgrevink PC, Lereim I. Personality profile among symptomatic and recovered patients with neck sprain injury, measured by MCMI-I acutely and 6 months after car accidents. *J Psychosom Res.* 1997;42:357-367.
4. Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehnsive Meta-Analysis.* Englewood, NJ: Biostat Inc; 2005.
5. Brison RJ, Hartling L, Pickett W. A prospective study of acceleration-extension injuries following rear-end motor vehicle collisions. *J Musculoskeletal Pain.* 2000;8:97-113.
6. Cassidy JD, Carroll LJ, Cote P, Lemstra M, Berglund A, Nygren A. Effect of eliminating compensation for pain and suffering on the outcome of insurance claims for whiplash injury. *N Engl J Med.* 2000;342:1179-1186.
7. Cooper HM, Hedges L. *The Handbook of Research Synthesis.* 1st ed. New York, NY: Russell Sage Foundation; 1994.
8. Cote 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:E445-458.
9. Ferrari R, Kwan O, Russell AS, Pearce JM, Schrader H. The best approach to the problem of whiplash? One ticket to Lithuania, please. *Clin Exp Rheumatol.* 1999;17:321-326.
10. Financial Services Commission of Ontario. *Superintendent's Guideline No. 06/07.* Toronto, Canada: Financial Services Commission of Ontario (FSCO); 2003.
11. Freeman MD, Croft AC, Rossignol AM. "Whiplash associated disorders: redefining whiplash and its management" by the Quebec Task Force. A critical evaluation. *Spine*. 1998;23:1043-1049.
12. Glass GV. Primary, secondary, and meta-analysis of research. *Educ Res.* 1976;5:3-8.
13. Harder S, Veilleux M, Suissa S. The effect of socio-demographic and crash-related factors on the prognosis of whiplash. *J Clin Epidemiol.* 1998;51:377-384.
14. Hartling L, Brison RJ, Ardern C, Pickett W. Prognostic value of the Quebec Classification of Whiplash-Associated Disorders. *Spine*. 2001;26:36-41.
15. Hartling L, Pickett W, Brison RJ. Derivation of a clinical decision rule for whiplash associated disorders among individuals involved in rear-end collisions. *Accid Anal Prev.* 2002;34:531-539.
16. Hedges L, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Meth.* 1998;3:486-504.
17. Hendriks EJ, Scholten-Peeters GG, van der Windt DA, Neeleman-van der Steen CW, Oostendorp RA, Verhagen AP. Prognostic factors for poor recovery in acute whiplash patients. *Pain.* 2005;114:408-416. http://dx.doi.org/10.1016/j.pain.2005.01.006
18. Kasch H, Bach FW, Jensen TS. Handicap after acute whiplash injury: a 1-year prospective study of risk factors. *Neurology*. 2001;56:1637-1643.
19. Kivioja J, Jensen I, Lindgren U. Early coping strategies do not influence the prognosis after whiplash injuries. *Injury*. 2005;36:935-940. http://dx.doi.org/10.1016/j.injury.2004.09.038
20. Kyhlback M, Thierfelder T, Soderlund A. Prognostic factors in whiplash-associated disorders. *Int J Rehabil Res.* 2002;25:181-187.
21. Nachemson AL, Jonsson E. *Neck and back Pain: The Scientific Evidence of Causes, Diagnosis, and Treatment.* Philadelphia, PA: Lippincott, Williams & Wilkins; 2000.
22. Nederhand MJ, Ijzerman MJ, Hermens HJ, Turk DC, Zilvold G. Predictive value of fear avoidance in developing chronic neck pain disability: consequences for clinical decision making. *Arch*

Phys Med Rehabil. 2004;85:496-501.
23. Pobereskin LH. Whiplash following rear end collisions: a prospective cohort study. *J Neurol Neurosurg Psychiatry.* 2005;76:1146-1151. http://dx.doi.org/10.1136/jnnp.2004.049189
24. Quinlan KP, Annest JL, Myers B, Ryan G, Hill H. Neck strains and sprains among motor vehicle occupants-United States, 2000. *Accid Anal Prev.* 2004;36:21-27.
25. Radanov BP, Sturzenegger M. Predicting recovery from common whiplash. *Eur Neurol.* 1996;36:48-51.
26. Radanov BP, Sturzenegger M, Di Stefano G. Long-term outcome after whiplash injury. A 2-year follow-up considering features of injury mechanism and somatic, radiologic, and psychosocial findings. *Medicine (Baltimore).* 1995;74:281-297.
27. Rosenstiel AK, Keefe FJ. The use of coping strategies in chronic low back pain patients: relationship to patient characteristics and current adjustment. *Pain*. 1983;17:33-44.
28. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull.* 1979;86:638-641.
29. Scholten-Peeters GG, Verhagen AP, Bekkering GE, et al. Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies. *Pain*. 2003;104:303-322.
30. Soderlund A, Olerud C, Lindberg P. Acute whiplash-associated disorders (WAD): the effects of early mobilization and prognostic factors in long-term symptomatology. *Clin Rehabil.* 2000;14:457-467.
31. Spitzer WO, Skovron ML, Salmi LR, et al. Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining "whiplash" and its management. *Spine*. 1995;20:1S-73S.
32. Sterling M. A proposed new classification system for whiplash associated disorders--implications for assessment and management. *Man Ther.* 2004;9:60-70. http://dx.doi.org/10.1016/j.math.2004.01.006
33. Sterling M, Jull G, Vicenzino B, Kenardy J. Sensory hypersensitivity occurs soon after whiplash injury and is associated with poor recovery. *Pain*. 2003;104:509-517.
34. Sterling M, Jull G, Vicenzino B, Kenardy J, Darnell R. Physical and psychological factors predict outcome following whiplash injury. *Pain*. 2005;114:141-148. http://dx.doi.org/10.1016/j.pain.2004.12.005
35. Sterner Y, Toolanen G, Gerdle B, Hildingsson C. The incidence of whiplash trauma and the effects of different factors on recovery. *J Spinal Disord Tech.* 2003;16:195-199.
36. Sturzenegger M, Radanov BP, Di Stefano G. The effect of accident mechanisms and initial findings on the long-term course of whiplash injury. *J Neurol.* 1995;242:443-449.
37. Tenenbaum A, Rivano-Fischer M, Tjell C, Edblom M, Sunnerhagen KS. The Quebec classification and a new Swedish classification for whiplash-associated disorders in relation to life satisfaction in patients at high risk of chronic functional impairment and disability. *J Rehabil Med.* 2002;34:114-118.
38. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther.* 1991;14:409-415.
39. Vlaeyen JW, Geurts SM, Kole-Snijders AM, Schuerman JA, Groenman NH, van Eek H. What do chronic pain patients think of their pain? Towards a pain cognition questionnaire. *Br J Clin Psychol.* 1990;29 (Pt 4):383-394.
40. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*. 2000;85:317-332.

APPENDIX A

Sampling	Yes	Partly	No	Can't tell
1. Were sample characteristics clearly stated?				
2. Were the characteristics of the refusers stated and were differences between refusers/acceptors investigated?				
3. Was the source population described?				
4. Were the subjects recruited within a reasonably narrow time frame?				
Methodology				
5. Was the exposure to the prognostic factor(s) captured using valid and reliable instruments?				
6. Were the investigators who captured outcome blinded to the presence/absence of prognostic factors?				
7. Did follow-up occur at the same point postinjury for all subjects?				
8. If the patients received intervention during the study, was it standardized, or was the effect of intervention statistically controlled for?				
9. Is the attrition rate acceptable?				
10. Is there evidence that subjects lost to follow-up were similar on baseline characteristics to those who completed the study?				
Analysis				
11. Are appropriate univariate crude estimates presented?				
12. Are appropriate multivariate analysis techniques employed?				
13. Is the sample size large enough for the number of variables investigated?				
14. Have the authors controlled for important confounders, either through stratification or statistical covariation?				
15. Was data manipulation appropriate?				
Results				
16. Were the results for prognostic factors presented in a clear and understandable fashion?				
17. Were the results for the main outcomes presented in a clear and understandable fashion?				

APPENDIX B

STATISTICAL CONVERSION PROCEDURES FOR DETERMINING EFFECT SIZE

The size of the effect is the ratio of the magnitude of the point estimate (in this case, the odds ratio) to its variability, which allows for comparisons across different scales by forcing point estimates into a similar metric. This gives clinicians an estimate of the actual amount of risk modulation expected, given the presence or absence of the different risk factors. The pooled effect size was calculated for the various predictors in one of the following ways.

When results for categorical risk variables are presented as frequencies of occurrence/nonoccurrence,[3-5,13,16,17,22,33] a 2-by-2 table is constructed, in which the rows represent the subject's status on the predictor (positive/negative), and the columns represent the status on the outcome (recovered/not recovered). The table below provides a graphic representation, with boxes labeled A, B, C, and D, representing the number of subjects that fall into each category. To calculate the odds ratio, the formula (A/B) ÷ (C/D) is used.

		Status on outcome measure: Recovered	Status on outcome measure: Not recovered
Status on predictor	Positive	A	B
	Negative	C	D

A 2-by-2 contingency table for determining odds ratios.

In papers where only computed odds ratios were presented with 95% confidence limits,[5,13,32] these data were entered directly into the database.

Some authors[1,3,5,15,16,17,32] opted to categorize continuous data such as baseline pain intensity, age, or years of education and calculate odds ratios using these sometimes arbitrarily defined categories. In these cases, the approach described above using a 2-by-2 table applied.

Other authors[4,17,22,28,33] presented baseline results in continuous form, providing means and standard deviations of continuous baseline risk factors for each group of recovered and nonrecovered subjects. The continuous data were converted to allow pooling with binary data. We used 2 different approaches to this problem, which allowed us to conduct a sensitivity analysis between the approaches and ensure that our results were not a result of artifact from the conversion procedures. The approaches are described below.

Approach 1 (Dr Charlie Goldsmith, personal communication, July 2007) Meaningful cut point values were determined for continuous measures based on the available literature (ie, high versus low) or on the cut points of similar measures from other papers in this review. Two normal distribution curves for each continuous predictor were constructed, 1 for the recovered and 1 for the nonrecovered groups, based on means and standard deviations presented in the published paper. The *z* value of the meaningful cut point was then determined for each distribution, and an estimation of the number of subjects that lay below that cut point and the number of subjects that lay above that cut point was entered into the database for statistical pooling.

Approach 2[4] The standardized mean difference (Cohen's *d*) was calculated by dividing the difference in mean values between the 2 groups by the pooled standard deviation within groups:

$$d = (mean_1 - mean_2) \div SD_{within}, \text{ where}$$

$$SD_{within} = \sqrt{[(N_1 - 1) \times SD_1^2 + (N_2 - 1) \times SD_2^2] \div [N_1 + N_2 - 2]}$$

The standard error of d was calculated as:

$$SE_d = \sqrt{([N_1 + N_2] \div [N_1 \times N_2]) + [d^2 \div 2 \times (N_1 + N_2)]}$$

Subsequently, the log odds ratio and SE were then calculated as:

$$\text{Log odds ratio} = \frac{\pi \times d}{\sqrt{3}}$$

$$SE_{\text{Log odds ratio}} = \sqrt{\pi^2 \times SE_d^2 \div 3}$$

As a sensitivity analysis, individual point estimates were compared across methods to determine if the odds ratio obtained through each approach was encompassed within the 95% confidence interval from the other approach. We then included the results from each approach in the database to estimate the pooled odds ratio for that predictor, and determined whether the difference in pooled odds ratio, when using the results from the 2 different approaches, changed the overall clinical picture for that predictor. The differences between the results obtained using the 2 approaches were insignificant for each instance in which this conversion was necessary. Therefore, the results from the dichotomization procedure have been reported, as they are more easily interpreted.

Data presented by Hendriks et al[17] presented a unique challenge for conversion as only univariate regression coefficients (beta values) with *P* values were provided. In this case the regression coefficients were input directly into the database as log odds ratios, and converted to the odds ratio. The 95% confidence interval was calculated using a *z* transformation, given beta coefficient (log odds ratio), and *P* value only (Mr Larry Stitt, personal communication):

$$z = \text{Log odds ratio} \div SE_{\text{Log odds ratio}}$$

This means that if the log odds ratio (beta coefficient) and *z* score are known, the equation can be rearranged to solve for SE:

$$SE_{\text{Log odds ratio}} = \text{Log odds ratio} \div z$$

With 1 degree of freedom, *z* is equal to $\sqrt{\chi^2}$. The χ^2 value was determined from a table of critical values, given the *P* value with 1 degree of freedom. To get the values for SE into the same metric as the log odds ratio, the natural log of the χ^2 was used. The formula for SE was then estimated as:

$$SE_{\text{Log odds ratio}} = \text{Log odds ratio} \div \sqrt{\ln(\chi^2)}$$