PAIN® 153 (2012) 1727–1734

PAIN®

www.elsevier.com/locate/pain

Assessment and validation of prognostic models for poor functional recovery 12 months after whiplash injury: A multicentre inception cohort study

Michele Sterling [a,*], Joan Hendrikz [a], Justin Kenardy [a], Eythor Kristjansson [b], Jean-Pierre Dumas [c], Ken Niere [d], Julie Cote [e,f], Sophie deSerres [f,g], Karine Rivest [f,g], Gwendolen Jull [h]

[a] Centre for National Research on Disability and Rehabilitation Medicine (CONROD), The University of Queensland, Australia
[b] Landspitalinn University Hospital, Reykjavik, Iceland
[c] School of Rehabilitation, University of Sherbrooke, Sherbrooke, Canada
[d] LaTrobe University, Melbourne, Victoria, Australia
[e] Department of Kinesiology and Physical Education, McGill University, Montreal, Canada
[f] Feil & Oberfeld/CRIR Research Centre, Jewish Rehabilitation Hospital, Laval, Canada
[g] School of Physical and Occupational Therapy, McGill University, Montreal, Canada
[h] CCRE Spinal Pain, Injury and Health, Division of Physiotherapy, The University of Queensland, Australia

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

ARTICLE INFO

Article history:
Received 23 January 2012
Received in revised form 13 April 2012
Accepted 2 May 2012

Keywords:
Whiplash-associated disorder
Prediction
Inception cohort
Validation

ABSTRACT

Uncertainty surrounds prognostic factors after whiplash injury. Previously we identified a prognostic model for 6-month pain-related disability in a cohort of 80 participants with acute whiplash. Predictors included initial disability, older age, decreased cold pain thresholds, decreased neck rotation movement, posttraumatic stress symptoms and decreased sympathetic vasoconstriction. The objective of this study was to externally validate this model. In a multicentre inception cohort study, 286 participants with acute whiplash (I, II or III) were assessed at <3 weeks and 12 months after injury. The Neck Disability Index (NDI) was the outcome. Observed and predicted NDI scores were generated using the published equation of the original model. Model discrimination between participants with no or mild disability from those with moderate to severe disability was examined by receiver operating characteristic curves. Initial NDI and cold pain threshold predicted current observed 12-month NDI scores (r^2 = 0.50, 95% confidence interval 0.42 to 0.58). There was a significant site effect, and the estimated marginal mean ± SE of 12-month NDI for Iceland (27.6 ± 1.79%) was higher than the other 3 sites (Melbourne 11.2 ± 5.03%, Canada 16.4 ± 2.36%, Brisbane 16.8 ± 1.17%). After adjusting for site, age and Impact of Events Scale scores regained significance (r^2 = 0.56, 95% confidence interval 0.48 to 0.64). The tested model was not precise in predicting NDI as a continuous variable. However, it found good accuracy to discriminate participants with moderate to severe disability at 12 months (area under the receiver operating characteristic curve 0.89 [95% confidence interval 0.84–0.94], $P < .001$) which is clinically useful.

1. Introduction

Whiplash injury after a motor vehicle crash (MVC) incurs enormous economic, social and personal costs. Recent data indicate that up to 50% of injured people will fail to fully recover [6]. Most recovery, if it occurs, will take place in the first 2–3 months after injury, after which time the condition plateaus [14,23]. Furthermore, the condition is resistant to most conservative interventions in both its acute [29] and chronic stages [13,28]. In view of these findings, it is important to consider prognostic factors that may be identified in the early acute stage of the condition so interventions may be specifically directed, potentially averting the course to chronicity.

Considerable uncertainty surrounds prognostic factors after whiplash injury. Whilst there are now 5 systematic reviews which have aimed to synthesise the literature in this area, all have noted the generally poor quality of many of the primary cohort studies [8,14,19,31,32]. Initially higher levels of pain reported soon after the MVC is the only consistently recognised factor that predicts poor functional recovery [8,14,19,31,32]. Other factors such as neck movement loss, hyperalgesic sensory responses and some psychological factors have shown inconsistent predictive capacity in some cohorts [31]. Whilst several studies have developed a predictive model, to our knowledge, there are no studies that have

* Corresponding author. Tel.: +61 7 3365 5344; fax: +61 7 3346 4603.
E-mail address: m.sterling@uq.edu.au (M. Sterling).

0304-3959/$36.00
http://dx.doi.org/10.1016/j.pain.2012.05.004

attempted to externally validate these models. Predictive models commonly perform poorly when applied to other patients and model validation is a necessity in order to have clinical utility [3]. External validation has been rarely performed in research of musculoskeletal pain including whiplash.

Previously we identified a prognostic model for the pain-related disability in an acute whiplash-associated disorders (WAD) cohort. Identified predictors included initial levels of pain-related disability, older age, decreased cold pain thresholds (cold hyperalgesia), decreased left neck rotation movement, posttraumatic stress symptoms and decreased sympathetic vasoconstriction [25]. The model explained 63% of the variance in pain and disability levels (Neck Disability Index, NDI [30]) at 6 months after MVC. Whilst the model was consistent with previous cohort studies in that initial higher levels of pain-related disability was a significant predictor, the additional variables almost doubled the rate of successful classification of participants who developed chronic moderate to severe symptoms [25]. Cold hyperalgesia and sympathetic disturbances may reflect altered central nervous system nociceptive processing [21] or, as shown in animal studies, stress system responses [2]. The use of approaches to modulate these changes may be an effective treatment strategy [17]. Preliminary data indicate that targeting posttraumatic stress symptoms in chronic WAD demonstrates the potential to decrease both posttraumatic stress and pain-related disability [10]. Thus, our initial model holds promise for the early evaluation of whiplash injury, but before it can be readily taken up, external validation is necessary.

We conducted a multicentre, international cohort study with the primary aim to externally validate a previously developed predictive model for poor functional recovery after whiplash injury.

2. Methods

2.1. Study design

The study was a multicentre prospective longitudinal study with an inception cohort of 286 people with acute whiplash injury (<3 weeks duration) after a MVC with follow-up at 12 months after injury. Participants were recruited at 4 centres: Brisbane, Australia; Melbourne, Australia; Montreal, Canada; and Reykjavik, Iceland. At all sites, the participants attended a university research laboratory for assessment at both assessment points. The study took place from 2005 to 2008. Ethical approval was obtained from the institutional medical ethics committee at each site. Informed consent was obtained from each participant.

2.2. Participants

Participants were included if they met the Quebec Task Force Classification of whiplash grades I, II or III [22] and were excluded if they were WAD IV (fracture or dislocation), if they experienced concussion, loss of consciousness or head injury as a result of the accident, and if they reported a history of whiplash, neck pain or headaches that required treatment. At 3 sites (Brisbane, Melbourne and Montreal), consecutive participants were recruited from primary care practices (medical and physiotherapy) and accident and emergency departments of local hospitals, and through general advertisement. These varied methods of recruitment were purposively used in order to reflect the nature of whiplash injury, where people may seek management from a variety of sources as well as not seek management at all. For logistical reasons, in Iceland, participants were recruited through the emergency department at Landspitali University Hospital in Reykjavik, Iceland, which maintains a database for people exposed to MVCs.

2.3. Outcome measure and predictor variables

The outcome was neck pain–related disability measured with the NDI, a valid, reliable and responsive measure [16]. There are 6 potential responses for each of 10 items ranging from no disability (0) to total disability (5). The overall score (out of 100) is calculated by totalling the responses of each individual item and multiplying by 2. The score was dichotomised into 2 diagnostic groups (0–28, mild or no disability; 30–100, moderate to severe disability) to compare diagnostic efficacy using the predictive models being examined [16].

Variables included in prognostic model for the validation analyses were measured at baseline and included initial NDI score, age at last birthday, range of cervical rotation to the left (ROML), cold pain threshold (CPT), a measure of sympathetic vasoconstriction (QI, quotient of integrals) and total score of the Impact of Events Scale (IES total).

Cervical range of movement was measured using an electromagnetic, motion-tracking device (Fastrak; Polhemus, Colchester, VT) according to previously established methodology [9]. ROML was used as the predictor variable [25]. Cold pain thresholds were measured over the mid cervical spine using the Thermotest system (Somedic, Farsta, Sweden). Triplicate recordings were taken at each site and the mean values used for analysis [27]. Peripheral vasoconstriction was measured over the skin of the fingertips of both hands using laser Doppler flowmetry (floLAB Monitor; Moor Instruments, Devon, England, UK) [26]. A provocation maneuver (inspiratory gasp), which is known to cause a short sympathetic reaction and cutaneous vasoconstriction, was performed [20]. Two quotients were calculated: the SRF parameter (sympathetic reflex), which represents the relative drop in the curve after provocation, and the quotient of integrals (QI), which also takes into account the duration of perfusion decrease [20]. A high QI and low SRF are indicative of an impaired vasoconstrictor response. QI was used in the analyses because it was a significant predictor variable in the initial developmental study [25]. The Impact of Events Scale (IES) is a 15-item questionnaire that measures current subjective stress related to a specific life event [11].

Assessors were blind to the results of the NDI questionnaire of each participant when they measured the predictor variables.

2.4. Sample size

Part of the validation process is to examine any bias and discrimination with regards to moderate to severe levels of pain and disability. Previous findings indicate that approximately 30% of those injured will develop chronic moderate to severe disability [25]. A sample size of 228 participants is required to be 95% confident that the true chronic moderate to severe disability is at least 0.25. Taking into account a 20% attrition rate, the recruitment target was 285 participants.

2.5. Analyses

Two predictive models for NDI at 12 months were examined and compared: (i) the model to be validated (Vmodel), which was previously developed for NDI at 6 months and reported by Sterling et al. [25]; and (ii) a regression model (Rmodel) developed with the current data, using the same set of 6 baseline variables used in the Vmodel, namely initial NDI, age last birthday, ROML, CPT, QI and IES total score. The current observed data was collected from several national and international sites. Thus, site was also investigated and taken into consideration in the analyses. As an initial step model differences in the parameter estimates, standard errors and statistical significance of the variables were examined and compared.

The validity of the prognostic Vmodel was evaluated using methods described by Altman and colleagues [1]. The fit of this model to the current data was initially assessed by comparing current observed NDI values at 12 months with predicted values generated using the published equation (Vmodel) and calculating subject differences. Mean differences were analysed by a paired-sample *t* test or the Wilcoxon signed ranks test, as appropriate, to test null hypotheses of no mean difference. Plots of differences (Vmodel) or residuals (Rmodel) against observed NDI at 12 months were examined. The association of positive and negative differences with a severity dichotomy of observed 12-month NDI (no or mild disability; moderate to severe disability) was described and tested by Pearson's chi-square test.

To determine how well the current observed NDI values at 12 months were calibrated using values calculated from the Vmodel, a scatter plot was produced with ideal and actual calibration lines superimposed.

The efficacy of calculated or predicted 12-month NDI values from the Vmodel and Rmodels to discriminate between subjects who were observed to have no or mild disability from those with moderate to severe disability was examined using receiver operating characteristic curves, sensitivity and specificity. Using the method described by Hosmer and Lemshow [12], sensitivity versus 1-specificity is plotted on a graph for each value calculated or predicted by the models, if used as a cutoff or criterion for disability diagnosis. Areas under the curve range between 0.5 (equal probability of correct diagnosis either way whatever criterion is chosen) and 1 (perfect discrimination). Furthermore, values of 12-month NDI, using either of the 2 models, can be read off from the data or graph and used for disability diagnosis corresponding to selected combinations of sensitivities and specificities.

Apart from QI (7% missing), there was less than 5% ($n = 14$) missing data for each of the remaining independent variables. None of the missing patterns for these variables was significantly linked to the outcome: NDI at 12 months. A total of 253 participants (88%) had complete data for all 6 independent variables. A missing value analysis indicated that those who had missing NDI at 12 months ($n = 29$) were significantly younger on average by 6 years (mean ± standard deviation [SD] age: 29.9 ± 8.6 years compared with 35.9 ± 13.4 years, $t = -3.3$, df = 45, $P = .002$), and their CPT was significantly higher by 3.7 degrees (mean CPT 17.7 compared with 14.0, $t = 2.4$, df = 270, $P = .013$). However, because these means and their differences are not considered to be clinically important, and because 257 (90%) had complete data for the outcome, listwise deletion was used, resulting in a sample of 225 participants.

SPSS Software (PASW V. 17) was used to do the statistical analysis, and Excel 2003 was used to produce the graphs. A *P* value of <.05 was deemed to indicate statistical significance.

3. Results

3.1. Participant recruitment and characteristics

Fig. 1 shows the flow of participants through the study at each site. Of the 286 participants who were recruited to the study, 257 (90%) completed the final 12-month assessment. A total of 225 participants (79%) had complete data for the regression analyses.

Participant age at last birthday in the collected data was (mean ± SD) 35.3 ± 13.08 years; 37.4% were men, and 29% reported moderate to severe disability. The number of participants recruited were as follows: Brisbane, Australia, 163 (57%); Melbourne, Australia, 13 (4.5%); Canada, 38 (13%); and Iceland 72 (25%). Table 1 shows the baseline characteristics of the participants at each site. This study did not aim to investigate the effect of treatment, and participants were free to pursue any form of treatment. The types and numbers of treatments received (including medication) were similar between the sites. In Brisbane, 41% of participants received treatment, 48% in Melbourne, 52% in Canada and 52% in Iceland. Physiotherapy was the most common form of treatment (Brisbane 25%, Melbourne 26%, Canada 26.3% and Iceland 17%). Other treatments received included chiropractic, acupuncture and massage. Medication usage was also similar between sites, with simple analgesics (Brisbane 32%, Melbourne 30%, Canada 29%, Iceland 31%) and nonsteroidal anti-inflammatory drugs (Brisbane 38%, Melbourne 39%, Canada 26%, Iceland 50%) the most common. Opioid-based medication (Brisbane 20%, Melbourne 18%, Canada 15%, Iceland 5%) and adjuvant medications (Brisbane 9%, Melbourne 11%, Canada 15%, Iceland 5%) were also used.

3.2. Prognostic model comparison

In the regression analysis, while initial pain-related disability and CPT were retained in the model (Rmodel), 4 of the 6 variables—Age, ROML, QI and IES total—which were significant in the Vmodel, did not have a significant effect on the current observed 12-month NDI outcome (Table 2). Furthermore, although ROML was nonsignificant in the Rmodel, the direction of the effect changed from the meaningful negative effect in the Vmodel to positive in the Rmodel. To address the fact that the data was collected at different sites, a site effect was added to the Rmodel which was reanalysed as a general linear model with covariates and site as a fixed effect. Type III sums of squares were examined because of unequal representation at the sites. The site effect was significant and the estimated marginal mean NDI at 12-months for Iceland (estimated marginal mean = 27.6, SE = 1.79) was significantly higher ($P < .05$ using Sidak's adjustment for multiple comparisons) than each of the other 3 site means, which in turn were not significantly different to each other (estimated marginal means [SE]: Melbourne, 11.2 [5.03]; Canada, 16.4 [2.36]; Queensland, 16.8 [1.17]). After adjusting for site differences in the Rmodel, Age and IES total regained significance, and although ROML remained nonsignificant, the parameter became meaningfully negative. These results seem to suggest that the Vmodel is unsuitable when data are collected from different centres and may not be applicable to outcomes collected after 6 months.

3.3. Model validity

Regression models are optimum in terms of fit inasmuch as the mean difference between observed and predicted values is always zero. However, the mean calculated value for the 12-month NDI using the Vmodel (mean ± SD, 32.2 ± 17.15) was significantly elevated ($t = 12.97$, df = 224, $P < .001$) compared with that for the observed data (19.2 ± 18.33), the mean difference being 13.0 ± 15.05 (Table 3).

The distribution of the differences, or fit, between the observed 12-month NDI and calculated values from the Vmodel were plotted against the observed values (Fig. 2). Most of the differences (83%) were negative, indicating that the majority of Vmodel predicted values overestimate the observed values. If the predictions were randomly distributed then half would be negative and half positive. The chi-square test for goodness-of -fit to this hypothesis was significant ($\chi^2 = 98.67$, df = 1, $P < .001$) indicating that the differences were not random. In contrast the rate of positive and negative residuals using the 2 Rmodels (46% and 47% positive respectively) were random (χ^2 goodness-of-fit tests, 1 df, $\chi^2 = 1.604$, $P = .205$ and $\chi^2 = 1.0$, $P = .317$ respectively).

A link between the observed 12-month NDI and the positive/negative differences (observed values minus model values) was investigated by examining cross-tabulations in which the observed

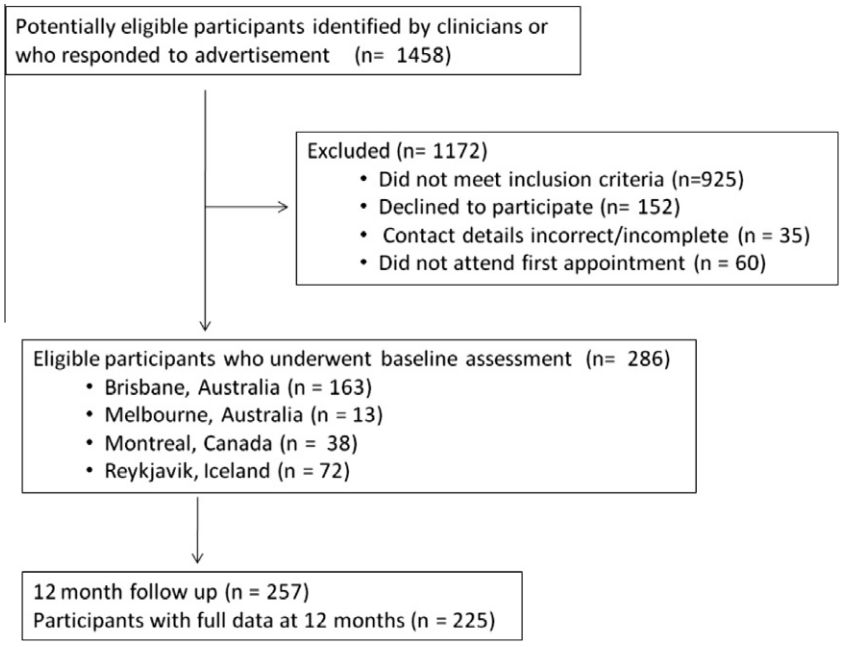

Fig. 1. Flow of participants through the study.

Table 1
Baseline characteristics of eligible participants.[a]

Variable	Brisbane, Qld, *N* = 163 (57%)	Melbourne, Vic, *N* = 13 (4.5%)	Montreal, Canada, *N* = 38 (13%)	Reykjavik, Iceland, *N* = 72 (25%)	*P*
Age (y)	37.3 ± 13.6	31.1 ± 13.4	34.3 ± 12.3	31.8 ± 11.5	$F = 3.51$; $P = .016$
Female	103 (64.8%)	11 (91.7%)	21 (55.3%)	41 (56.9%)	$\chi^2 = 7.534$; $P = .057$
MVC					$\chi^2 = 7.59$; $P = .576$
Rear end	67 (41.1%)	7 (53.8%)	16 (42.1%)	32 (44.4%)	
Front end	52 (31.9%)	1 (7.7%)	8 (21.1%)	18 (25.0%)	
Both rear and front	22 (13.5%)	3 (23.1%)	5 (13.2%)	11 (15.3%)	
Side impact	22 (13.5%)	2 (15.4%)	9 (23.7%)	11 (15.3%)	
Filed claim for compensation during study period	89 (55.6%)	3 (23.1%)	12 (31.6%)	33 (45.8%)	$\chi^2 = 8.692$; $P = .034$
Postsecondary education	48 (47.5%)	8 (57.1%)	21 (51.2%)	29 (38.2%)	$\chi^2 = 3.136$; $P = .371$
Changed work status as result of injury	42 (27.1%)	2 (15.4%)	13 (35.1%)	15 (22.7%)	$\chi^2 = 2.784$; $P = .426$
Neck pain intensity (VAS)	3.7 ± 2.02	4.2 ± 1.78	5.5 ± 2.11	4.7 ± 1.92	$F = 10.78$; $P < .001$
Initial disability (NDI)	33.2 ± 18.4	28.6 ± 15.7	32.6 ± 16.9	30.5 ± 15.8	$F = 0.58$; $P = .63$
Initial QI	55.8 ± 23.6	53.56 ± 14.4	94.0 ± 34.2	71.7 ± 19.6	$F = 27.76$; $P < .001$
Initial CPT	14.5 ± 8.0	11.2 ± 8.9	16.4 ± 6.5	13.4 ± 7.4	$F = 1.93$; $P = .125$
Initial ROML	53.5 ± 18.6	53.9 ± 10.2	55.3 ± 16.2	69.9 ± 13.2	$F = 15.34$; $P < .001$
Initial IES total	23.1 ± 16.7	19.2 ± 12.1	18.7 ± 17.6	14.2 ± 16.6	$F = 4.68$; $P = .003$

MVC, motor vehicle crash; VAS, visual analogue scale; NDI, Neck Disability Index; QI, quotient of integrals; CPT, cold pain threshold; ROML, range of cervical rotation to the left; IES, Impact of Events Scale.

[a] Data are presented as *n* (%) or mean ± SD.

Table 2
Parameter estimates from models with site (*n* = 225).

Predictor	Vmodel		Rmodel		Rmodel with site	
	Coefficient (SE)	*P*	Coefficient (SE)	*P*	Coefficient (SE)	*P*
Adjusted R^2	0.50		0.50		0.56	
Constant	11.74 (10.89)	.285	−12.65 (6.221)	.043	−6.09	.314
Site compared						
Brisbane	NA		NA			
Melbourne	NA		NA		−5.52	.284
Canada	NA		NA		−0.39	.889
Iceland	NA		NA		10.81	<.001
NDI	0.387 (0.083)	<.001	0.652 (0.068)	<.001	0.587 (0.065)	<.001
Age	0.387 (0.108)	.001	0.127 (0.066)	.055	0.135 (0.063)	.033
ROML	−0.178 (0.106)	.050	0.044 (0.062)	.481	−0.071 (0.063)	.257
CPT	0.505 (0.199)	.010	0.328 (0.133)	.014	0.302 (0.126)	.017
QI	−0.147 (0.070)	.040	−0.031 (0.032)	.334	−0.049 (0.036)	.174
IES total	0.338 (0.094)	.006	0.070 (0.059)	.234	0.113 (0.056)	.046

Vmodel, model to be validated; Rmodel, regression model; NDI, Neck Disability Index; ROML, range of cervical rotation to the left; CPT, cold pain threshold; QI, quotient of integrals; IES, Impact of Events Scale.

Table 3
Descriptive statistics for the 225 observed means and those predicted using the 3 models.

Variable	Observed NDI at 12 mo	Calculated NDI from Vmodel	Predicted NDI from Rmodel	Predicted NDI from Rmodel including site
SD	18.326	17.145	13.171	13.894
Minimum	0	−0.22	−5.36	−5.37
Maximum	80	84.39	54.49	55.33
25th Percentile	4	19.41	8.81	8.25
50th Percentile	14	29.82	16.27	16.97
75th Percentile	32	42.44	27.64	27.11

NDI, Neck Disability Index; Vmodel, model to be validated; Rmodel, regression model.

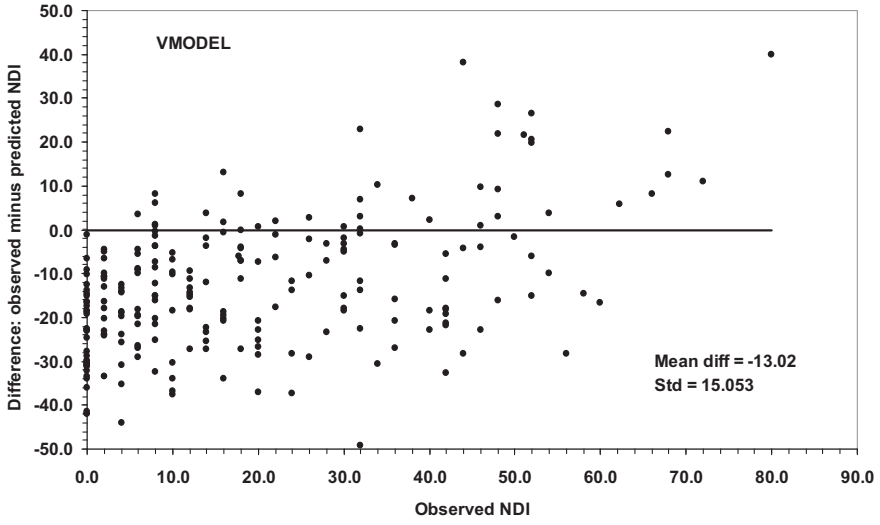

Fig. 2. Accuracy of prediction of NDI at 12 months using the model developed and published for 6-month data in 2005 (Vmodel).

Table 4
Distribution of the differences between the observed values and the estimated values using the Vmodel.

Difference between observed and predicted NDI with the Vmodel	Observed NDI at 12 mo		Total
	None or mild	Moderate to severe	
<0 (Vmodel over estimates)	147 (78.6%)	40 (21.4%)	187 (83.1%)
⩾0 (Vmodel under estimates)	12 (31.6%)	26 (68.4%)	38 (16.9%)
Total	159 (70.7%)	66 (29.3%)	225

Vmodel, model to be validated; NDI, Neck Disability Index.

12-month values were dichotomised into 2 severity groups of interest: No or mild disability (NDI) and moderate to severe disability (NDI) (Table 4). For all models (Vmodel and Rmodels) the distribution of positive and negative differences, or residuals, was not independent of the severity of value observed. Over 78% of the negative differences or residuals (over-estimates) were significantly linked to no or only mild observed disability (Pearson's $\chi^2 > 33.0$, 1 df, $P < .001$, for all models). These results were further evidence that the differences, or residuals, did not have a random distribution. Thus, none of the models seems to fit particularly well.

3.4. Calibration

The accuracy of calibration of an NDI value at 12 months using values calculated by the Vmodel is illustrated in a scatter plot of observed 12-month NDI plotted against calculated NDI (Fig. 3). A value calculated from the parameter coefficients of the 6 predictors in the retrospective Vmodel would first have to be multiplied by 0.69 and then have 2.9 subtracted before it could be used as a proxy for NDI at 12 months. However, even then, this is not an exact prediction because it is subject to the random variation about the prediction line (59%) because only 41% (R^2) of the calibration is determined by this equation.

3.5. Discrimination

The receiver operating characteristic curve analyses indicated that any of the models (Vmodel, Rmodel or Rmodel adjusted for site) could be used to discriminate those with no or with mild disability from those with moderate to severe disability levels, at 12 months (areas under the receiver operating characteristic curve: 0.85 [0.79–0.91], 0.89 [0.84–0.94], and 0.91 [95% confidence interval 0.86–0.95] respectively, all $P < .001$). This means that even if there is disparity between observed values and values calculated or predicted by the models, a value for the latter can still be chosen as the cutoff value for predicting severity of NDI at 12 months with specified sensitivity and/or specificity. Table 5 illustrates the cut-offs, specificity and positive and negative predictive values (PPV and NPV) corresponding to sensitivities of 80% and 90%, for each

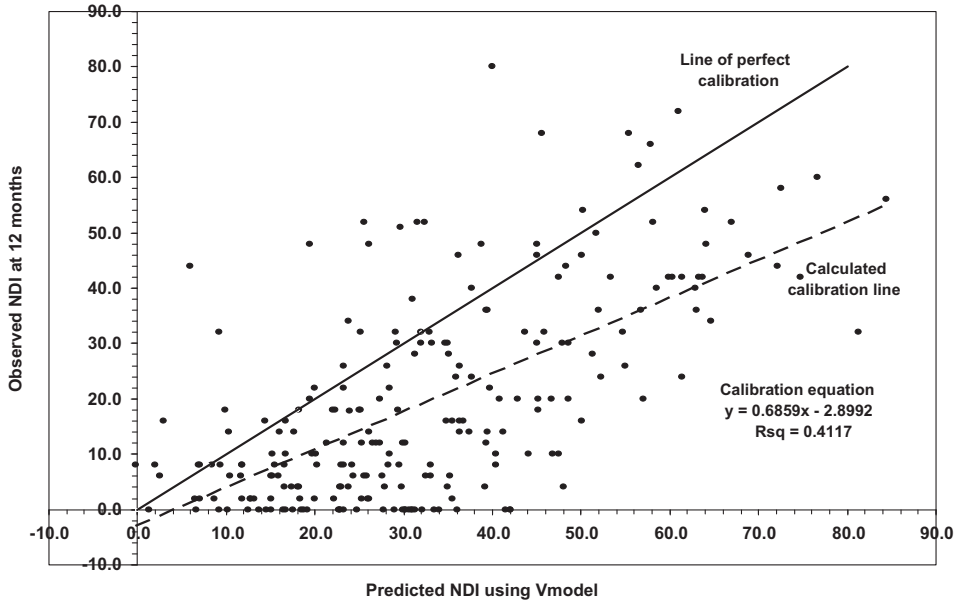

Fig. 3. Calibration of NDI at 12 months using the Vmodel.

Table 5
Sensitivity (90% and 80%) to moderate to severe NDI outcome at 12 months, and specificity to none–mild outcome corresponding to cut values generated from the predictive models.[a]

Sensitivity	Vmodel	Rmodel	Rmodel (with site)
Sensitivity 90%			
Cut point	29.0	16.93	18.2
Specificity	63%	69%	70%
PPV	50%	54%	55%
NPV	93%	94%	94%
Sensitivity 80%			
Cut point	32.0	21.7	22.5
Specificity	71%	79%	81%
PPV	53%	62%	64%
NPV	90%	91%	91%

NDI, Neck Disability Index; Vmodel, model to be validated; Rmodel, regression model; PPV, positive predictive value; NPV, NPV.

[a] Discrimination is for predicted values at or above the cut value; cut point is for the highest specificity corresponding to the sensitivity.

model. The regression model could be used anywhere as location or site has a minimal effect on discrimination efficacy and cut value. On the other hand the Vmodel is less specific to those with none or mild NDI and it decreases as sensitivity increases. Low PPV percents indicate that none of the models are particularly reliable in terms of accurately predicting moderate to severe NDI outcomes.

4. Discussion

Chronic pain and disability are common occurrences after whiplash injury [6], and the capacity to predict those who will follow such a clinical pathway will be important. To our knowledge, this is the first study to attempt external validation of a predictive model for poor functional recovery after whiplash injury. Whilst the model was not precise in terms of predicting NDI scores as a continuous variable, it showed good accuracy to discriminate participants with moderate to severe disability at 12 months from recovered participants and those with only mild disability. This latter finding is clinically relevant.

The tested model was not a good fit when external validation was attempted. The model tended to overestimate predicted levels of disability (NDI) at 12 months, and this was more so for those who would recover or eventually report lower disability. Moreover only 2 of the 6 variables (initial disability levels, cold pain threshold) in the predictive set were significant predictors. Prognostic models usually perform less well when tested in new cohorts [3]. This may be due to the sample size of the original cohort, differences in the patient setting of the new cohort including differences in measurement methods, health care systems and patient characteristics [3]. Our results indicate that these latter factors played a role in the poor calibration of our model, where it was demonstrated that the prognostic model performed differently depending upon test site. Whilst the Australian and Canadian sites were similar, in Iceland, the estimated disability levels were significantly greater by 11–17%, a clinically relevant difference for the NDI [16]. When site was taken into account in the analyses, the model showed improved fit, with additional variables (age, posttraumatic stress symptoms) becoming significant predictors.

The reasons for greater predicted pain-related disability in Iceland are not clear. The personnel involved in the trial were highly experienced in the use of the study tools, so measurement error seems unlikely. It is well documented that the experience of pain differs between ethnocultural groups in conditions such as arthritis and headache [5] and this likely occurs for whiplash as well [18]. The Icelandic participants were slightly younger by 5 1/2 years on average than the Brisbane participants but this gap closed between them and the other 2 sites. There was no significant difference in gender distribution between the sites. There was significant variation in average baseline levels of pain, where Icelandic and Canadian participants reported higher levels compared to their Australian counterparts, but there was no difference in disability levels between sites. There was no difference between sites in the nature of the MVC in terms of direction of impact. It is possible that environmental factors such as health and/or compensation systems between the sites may be important. There was a significant difference in the percentage of participants at each site who submitted a compensation claim (Brisbane 55.6%, Melbourne 23%, Canada 31.6%, Iceland 45.85%). However, these differences would not seem to be relevant because the lowest percentages of claimants were in Melbourne and Canada—sites were not signifi-

cantly different from Brisbane (which had the highest number of claimants) in outcome. Different insurance schemes between sites also do not seem relevant to our findings, where Iceland and Brisbane have fault-based schemes with Montreal, Canada, and Melbourne, Victoria, having no-fault-based schemes. The only clear difference is that Icelandic participants were recruited only from accident and emergency departments, whereas wider recruitment from primary care and the general community was conducted at the other sites. It is possible that people attending emergency departments report higher levels of pain and disability; however, contrary to this suggestion, observed site differences in other aspects are not unique to and even favour Iceland. For example, Iceland presented the highest mean ROML and lowest mean IES score. Nevertheless, our findings indicate that the prognostic model cannot be automatically taken up in every jurisdiction.

Although sensitivity and specificity are unaffected by study prevalence rates of moderate to severe disability, PPV and NPV are affected if the observed prevalence rate is different to that in the target population. The prevalence rate for moderate to severe disability in this sample is 29%, consistent with international data [6,14]. Despite the poor calibration, the accuracy of the model was good at discriminating those with persistent moderate to severe disability. Table 5 provides cutoffs or criterion values for each of the 3 models for 80% and 90% sensitivity. For example using the site adjusted model at 80% sensitivity with cutoff NDI score = 22.5, the specificity is 81%; PPV is 64% and NPV 91%. This would indicate that using this criterion in the model is good at identifying both those at risk of developing moderate to severe symptoms and those not. Also, the majority (91%) predicted to be less severe by the model criterion are indeed so, but 36% of those who are classified as being at risk by the model could recover. This suggests that those identified as being at risk in the early injury stage using the model be monitored further to determine if recovery does occur. Thus, the model or models could be useful for initial screening purposes. Injured people who transition to chronic moderate to severe disability incur the majority of the costs; thus, it is particularly important to be able to identify these individuals. This could be argued to be a more clinically relevant approach than to attempt to predict the specific NDI score of individual patients. Thus, in clinical terms, the model will be useful in gauging a patient's propensity to develop a chronic condition.

To our knowledge, this is the first study to validate a predictive model for whiplash, so we have no other similar studies for comparison. Nonetheless, we can compare individual predictor variables. All of the variables in our model have demonstrated prognostic capacity in previous studies. Initial levels of pain [31] and disability [6] are the most consistent predictors of poor recovery, and our findings validate these as important in the development of chronic pain after whiplash. The early presence of cold hyperalgesia has been demonstrated to predict chronic moderate to severe disability and psychological trajectories after injury [24], and decreased cold tolerance predicted poor recovery in a cohort study conducted in Sweden [15]. Williamson and colleagues [32] identified posttraumatic stress symptoms as having some prognostic capacity, and later cohort studies would support this finding [4]. Older age and neck movement are inconsistent predictors [31] and our results would support this inconsistency, where age was only significant after adjustment for site differences and neck movement was not significant.

In many jurisdictions, acute WAD is managed in primary care, and clinicians require easily measured indicators of poor prognosis. Of the 6 variables comprising the model, initial NDI scores and CPT were the only significant variables before site adjustment was made with age and IES scores becoming significant after site adjustment. Left neck rotation range and the QI quotient of vasoconstriction were not significant in either model. This suggests that the variables of NDI, CPT, age and posttraumatic stress symptoms (IES scores) be used in the clinical evaluation of acute WAD to broadly identify those who will develop persistent moderate to severe disability. Of these factors, only CPT is difficult to measure clinically, but it may be possible via the use of thermorollers or even via the application of ice [7]. Additionally, most factors comprising the model are potentially modifiable and may assist in the development of management strategies directed at them. This requires further research.

The strengths of this study are that this is first investigation into the validation of using a prior exploratory predictive model on a new multicentre international inception cohort over a 12-month period with high rates of follow-up. A limitation of the study is that we did not have equal participant numbers at each site, as was our original aim. The reason for this discrepancy was recruitment difficulties in Melbourne and Canada. The research units in Brisbane and Iceland had greater experience in the recruitment of participants with acute WAD, with well-established recruitment channels. Additional participants were recruited from the Brisbane site to make up the shortfall. This factor may have contributed to the site difference demonstrated for the predictive model. It may also have created bias in the study sample, and this should be considered when interpreting the study's findings. Nevertheless, it is the first time that a cohort study has included participants from different countries and jurisdictions.

In summary, our previously identified predictive model did not demonstrate precision in predicting 12-month disability scores in a new multicentre inception cohort. However, the model has good discriminating value for differentiating participants with moderate to severe disability at 12 months, which is clinically useful. Most of the predictors in the model are potentially modifiable and as such suggest that strategies directed towards their modification may assist in preventing the development of chronicity. The influence of implementation of the model on health outcomes also requires further investigation.

Conflict of interest statement

The authors report no conflict of interest.

Acknowledgements

The study was funded by an unrestricted grant from the Australian Research Council (LP0560611), the Motor Accident Insurance Commission, Qld (MAIC), and Suncorp General Insurance, Qld. The site in Canada received funding from the Quebec Rehabilitation Research Network (REPAR). The site in Iceland received funding from the Icelandic Safety Research Council. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. MS received a research fellowship from the National Health and Medical Research Council (NHMRC), Australia. Her appointment was also funded from an unrestricted grant from the Motor Accident Insurance Commission (MAIC) to the University of Queensland. JK's appointment was funded from an unrestricted grant from the MAIC to The University of Queensland.

References

[1] Altman D, Vergouwe Y, Royston P, Moons K. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.
[2] Bardin L, Malfetes N, Newman-Tancredi A, Depoortere R. Chronic restraint stress induces mechanical and cold allodynia, and enhances inflammatory pain in rat: relevance to human stress-associated painful pathologies. Behav Brain Res 2009;205:360–6.

[3] Bland D, Vergouwe Y, Royston P, Moons K. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:1432–5.
[4] Buitenhuis J, Jaspers J, Fidler V. Can kinesiophobia predict the duration of neck symptoms in acute whiplash? Clin J Pain 2006;22:272–7.
[5] Campbell C, Edwards R, Fillingim R. Ethnic differences in responses to multiple experimental pain stimuli. Pain 2005;113:20–6.
[6] Carroll L, Holm L, Hogg-Johnson S, Cote P, Cassidy D, Haldeman S, Nordin M, Hurwitz E, Carragee E, van der Velde G, Peloso P, Guzman J. Course and prognostic factors for neck pain in whiplash-associated disorders (WAD). Results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. Spine 2008;33:583–92.
[7] Cathcart S, Pritchard D. Relability of pain threshold measurement in young adults. J Headache Pain 2006;7:21–6.
[8] Cote P, Cassidy D, Carroll L, Frank J, Bombardier C. A systematic review of the prognosis of acute whiplash and a new conceptual framework to synthesize the literature. Spine 2001;26:E445–58.
[9] Dall'Alba P, Sterling M, Trealeven J, Edwards S, Jull G. Cervical range of motion discriminates between asymptomatic and whiplash subjects. Spine 2001;26:2090–4.
[10] Dunne R, Kenardy J, Sterling M. A randomised controlled trial of cognitive behavioural therapy for the treatment of PTSD in the context of chronic whiplash. Clin J Pain 2011 [Epub ahead of print].
[11] Horowitz M, Wilner N, Alvarez W. Impact of event scale: a measure of subjective stress. Psychosom Med 1979;41:209–18.
[12] Hosmer D, Lemshow S. Applied logistic regression. 2nd ed. Chicheser, UK: John Wiley and Sons; 2000.
[13] Jull G, Sterling M, Kenardy J, Beller E. Does the presence of sensory hypersensitivity influence outcomes of physical rehabilitation for chronic whiplash? A preliminary RCT. Pain 2007;129:28–34.
[14] Kamper S, Rebbeck T, Maher C, McAuley J, Sterling M. Course and prognostic factors of whiplash: a systematic review and meta-analysis. Pain 2008;138:617–29.
[15] Kasch H, Qerama E, Bach F, Jensen T. Reduced cold pressor pain tolerance in non-recovered whiplash patients: a 1 year prospective study. Eur J Pain 2005;9:561–9.
[16] MacDermid J, Walton D, Avery S, Blanchard A, Etruw E, McAlpine C, Goldsmith C. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 2009;39:400–17.
[17] Nijs J, Van Oosterwijck J, De hertogh W. Rehabilitation of chronic whiplash: treatment of cervical dysfunctions or chronic pain syndrome? Clin Rheumatol 2009;28:243–51.
[18] Partheni M, Constantoyannis C, Ferrari R, Nikiforidis G, Voulgaris S, Papadakis N. A prospective study of the outcome of acute whiplash injury in Greece. Clin Exp Rheumatol 2000;18:67–70.
[19] Scholten-Peeters G, Verhagen A, Bekkering G, van der Windt D, Barnsley L, Oostendorp R. Prognostic factors of whiplash associated disorders: a systematic review of prospective cohort studies. Pain 2003;104:303–22.
[20] Schurmann M, Gradl G, Andress H, Furst H, Schildberg F. Assessment of peripheral sympathetic nervous system function for diagnosing early post-traumatic complex regional pain syndrome type I. Pain 1999;80:149–59.
[21] Seifert F, Maihofner C. Representation of cold allodynia in the human brain—a functional MRI study. Neuroimage 2007;35:1168–80.
[22] Spitzer W, Skovron M, Salmi L, Cassidy J, Duranceau J, Suissa S, Zeiss E. Scientific monograph of quebec task force on whiplash associated disorders: redefining 'whiplash' and its management. Spine 1995;20:1–73.
[23] Sterling M, Hendrikz J, Kenardy J. Developmental trajectories of pain/disability and PTSD symptoms following whiplash injury. Pain 2010;150:22–8.
[24] Sterling M, Hendrikz J, Kenardy J. Similar factors predict disability and PTSD trajectories following whiplash injury. Pain 2011;152:1272–8.
[25] Sterling M, Jull G, Vicenzino B, Kenardy J, Darnell R. Physical and psychological factors predict outcome following whiplash injury. Pain 2005;114:141–8.
[26] Sterling M, Kenardy J. The relationship between sensory and sympathetic nervous system changes and acute posttraumatic stress following whiplash injury—a prospective study. J Psychosom Res 2006;60:387–93.
[27] Sterling M, Pettiford C, Hodkinson E, Curatolo M. Psychological factors are related to some sensory pain thresholds but not nociceptive flexion reflex threshold in chronic whiplash. Clin J Pain 2008;24:124–30.
[28] Stewart M, Maher C, Refshauge K, Herbert R, Bogduk N, Nicholas M. Randomised controlled trial of exercise for chronic whiplash associated disorders. Pain 2007;128:59–68.
[29] Verhagen A, Scholten-Peeters G, van Wijngaarden S, de Bie R, Bierma-Zeinstra S. Conservative treatments for whiplash. Cochrane Database Syst Rev 2007;2:CD003338.
[30] Vernon H, Mior S. The neck disability index: a study of reliability and validity. J Manipulative Physiol Ther 1991;14:409–15.
[31] Walton D, Pretty J, MacDermid J, Teasell R. Risk factors for persistent problems following whiplash injury: results of a systematic review and meta-analysis. J Orthop Sports Phys Ther 2009;39:334–50.
[32] Williamson E, Williams M, Gates S, Lamb S. A systematic review of psychological factors and the development of late whiplash syndrome. Pain 2008;135:20–30.