Spine

SPINE Volume 44, Number 22, pp E1298–E1310

OPEN

Broad External Validation and Update of a Prediction Model for Persistent Neck Pain After 12 Weeks

Birgitte Lawaetz Myhrvold, MSc,[*] Alice Kongsted, PhD,[†,‡] Pernille Irgens, MSc,[*] Hilde Stendal Robinson, PhD,[*] Magne Thoresen, PhD,[§] and Nina Køpke Vøllestad, PhD[*]\

Study Design. A prospective observational study.

Objective. To externally validate the prediction model developed by Schellingerhout and colleagues predicting global perceived effect at 12 weeks in patients with neck pain and to update and internally validate the updated model.

Summary of Background Data. Only one prediction model for neck pain has undergone some external validation with good promise. However, the model needs testing in other populations before implementation in clinical practice.

Methods. Patients with neck pain (n = 773) consulting Norwegian chiropractors were followed for 12 weeks. Parameters from the original prediction model were applied to this sample for external validation. Subsequently, two random samples were drawn from the full study sample. One sample (n = 436) was used to update the model; by recalibration, removing noninformative covariates, and adding new possible predictors. The updated model was tested in the other sample (n = 303) using stepwise logistic regression analysis. Main outcomes for performance of models were discrimination and calibration plots.

Results. Three hundred seventy patients (47%) in the full study sample reported persistent pain at 12 weeks. The performance of the original model was poor, area under the receiver operating characteristics curve was 0.55 with a Confidence Interval of 0.51–0.59. The updated model included Radiating pain to shoulder and/or elbow, education level, physical activity, consultation-type (first- time, follow-up or maintenance consultation), expected course of neck pain, previous course of neck pain, number of pain sites, and the interaction term Physical activity##Number of pain sites. The area under the receiver operating characteristics curve was 0.65 with a 95% Confidence Interval of 0.58–0.71 for the updated model.

Conclusion. The predictive accuracy of the original model performed insufficiently in the sample of patients from Norwegian chiropractors and the model is therefore not recommended for that setting. Only one predictor from the original model was retained in the updated model, which demonstrated reasonable good performance predicting outcome at 12 weeks. Before considering clinical use, a new external validation is required.

Key words: chiropractic, external validation, neck pain, performance accuracy, prediction rule, prognosis, updating.

Level of Evidence: 3

Spine 2019;44:E1298–E1310

From the [*]Department of Interdisciplinary Health Sciences, Institute of Health and Society, University of Oslo, Oslo, Norway; [†]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark; [‡]Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark; and [§]Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway.

Acknowledgment date: February 25, 2019. First revision date: May 1, 2019. Acceptance date: May 27, 2019.

The manuscript submitted does not contain information about medical device(s)/drug(s).

The Norwegian Chiropractic Research Foundation ''Et liv i bevegelse (ELIB) and the European Chiropractic Union (ECU) funds were received in support of this work.

No relevant financial activities outside the submitted work.

Address correspondence and reprint requests to Birgitte Lawaetz Myhrvold, MSc, Department of Interdisciplinary Health Sciences, Institute of Health and Society, University of Oslo, P.O. Box 1089, Blindern, 0317 Oslo, Norway; E-mail: b.l.myhrvold@medisin.uio.no

DOI: 10.1097/BRS.0000000000003144

Neck pain is common with an annual prevalence of 30% to 50% in the general population,[1–5] and causes substantial disability and economic expenses.[1,3,6–8] Since the highest burden of disability and health expenses, relate to a small proportion of those with neck pain there is a need to identify factors that can predict patients' likely outcome and tailor management accordingly. Although a number of predictors associated with neck pain have been identified, evidence related to prediction of the clinical course of recovery and decisions regarding choice of treatment is limited and inconsistent.[6,9,10] A recent literature review concluded that the quality of studies developing prediction models on neck pain varies, and that the majority of models lack proper validation.[11] In 2010, Schellingerhout *et al*[12] identified predictive factors and developed

TABLE 1. Measurement of Patient and Clinician Reported Baseline Variables of the Original Sample (see [12]) and the Present Chiropractic Study Sample Divided into 5 Health Domains

	Measurement Original Sample	Coding in the Original Sample	Predictors Included in the Original Model	Measurement Present Study Sample	Coding in the Present Study Sample	Predictors Included in the Updated Model
Health domains/predictors						
Domain 1: Sociodemographic and physical factors						
Gender	Female/Male	Female = 1/Male = 0		Female/Male	Female = 1/Male = 0	
Age	Years	Continuous scale, 18–70	X	Years	Continuous scale, 18–70	
Education level	High/Medium/Low	High/Medium/Low		High/Medium/Low*	High = 1/Medium = 2/ Low = 3	X
Employment status	Employed/Not employed	Employed = 1/Not employed = 0	X	Employed/Not employed	Employed = 1/Not employed = 0	
Biking	-	-		Yes daily/No rarely	Yes = 1/No = 0	
Physical activity	-	-		Never/Less than once a week/Once a week/2–3 times a week/ more than 3 times a week	≥1 per week = 1/<1 per week = 0	X
Domain 2: Clinician reported						
Consultation-type	-	-		"First-time consultation" "Follow-up consultation" "Maintenance consultation"	First-time = 1 Follow-up = 2 Maintenance = 3	X
Domain 3: Neck pain history						
Previous neck complaints	Previous episode of neck complaints	Yes = 1/No = 0	X	No first time/Yes, 1–3 times previously/Yes, >3 times previously/Yes, more or less chronically neck pain	Yes = 1/No = 0	
Traumatic cause neck complaint	Cause of neck pain	Trauma = 1/No trauma = 0	X	Cause of neck pain	Trauma = 1/No trauma = 0	
Radiating pain to shoulder and/or elbow	Radiation of pain to elbow or shoulder	Yes = 1/No = 0	X	NPQ Last 7 days	Yes = 1/No = 0	X
Pain intensity	NRS Numeric rating scale, 11-point	0 = 0; 1–10 = 1	X	NRS Numeric rating scale, 11-point	0 = 0; 1–10 = 1	
Duration current episode	-	<1 month = 1/1–3 months = 2/>3 months = 3			<1 month = 1/1–3 months = 2/>3 months = 3	
NDI	NDI	Continuous scale 0–50		NDI	Continuous scale 0–50	
Use of painkillers	-	-		Yes = 1/No = 0	Yes = 1/No = 0	
Patient previous trajectory pattern	-	-		Five visual trajectory patterns of neck pain	Single episode = 1 Episodic pain = 2 Mild pain/ recovering = 3 Fluctuating pain = 4 Moderate/severe pain = 5 Neither/ Unsure = 9	X
Sick-listing for present neck pain	-	-		Yes = 1/no = 0	Yes = 1/no = 0	
Insurance claim of present neck pain	-	-		Yes = 1/no = 0	Yes = 1/no = 0	
Domain 4: General health/Comorbidity						
Low back pain	Accompanying low back pain	Yes = 1/No = 0	X	NPQ Last 7 days	Yes = 1/No = 0	
Headache	Accompanying headache	Yes = 1/No = 0	X	NPQ Last 7 days	Yes = 1/No = 0	
Number of pain sites	-	-		NPQ Sum-score, Last 7 days	Continuous scale, 0–10	X
Health status	Eq5d, Quality of life	100 mm VAS, 0–100	X	Eq5d, Quality of life	100 mm VAS, 0-100	
Eq5d-index	-	-		Index scale Full health has a value of 1 and dead a value of 0	Continuous scale, 0–1	
Domain 5: Psychological						
Kinesophobia	TAMPA scale	Continuous scale scale, 17–68		"In your normal daily activities, how much trouble do you have from your neck pain complaints?"	Continuous scale, 0-10	

TABLE 1 (*Continued*)

	Measurement Original Sample	Coding in the Original Sample	Predictors Included in the Original Model	Measurement Present Study Sample	Coding in the Present Study Sample	Predictors Included in the Updated Model
HSCL-score	-	-		HSCL-10 scale	Continuous scale,	
Örebro score	-	-		Örebro Screening Questionnaire Sum score ranged from 0 to 100	Continuous scale, 0–100	
Patient expected trajectory pattern	-	-		Five visual trajectory patterns of neck pain	Single episode = 1 Episodic pain = 2 Mild pain/ recovering = 3 Fluctuating pain = 4 Moderate/severe pain = 5 Neither/ Unsure = 9	X
Interaction terms						
Physical activity## Number of pain sites						X

*High = University College or university over 4 years, medium = University College or university under 4 years or upper secondary school/high school, low = vocational school or lower secondary school.

HSCL-10: measuring anxiety and depression,[16] TAMPA scale, Örebro Screening Questionnaire: predicting long-term disability and failure to return to work[17,18]; NPQ: Nordic pain questionnaire measuring musculoskeletal symptom prevalence,[19] NRS: the 11-point numerical rating scale,[20] NDI: the Neck Disability Index measuring disability,[21] Eq5d: Health-related quality of life.[22] Cells marked with a dash (-) indicate the predictor is NOT in the original model.

a prediction model aiming at identifying patients at risk of persistent neck pain complaints after 6 months and confirmed its validity in a separate patient sample (external validation). To our knowledge, this is the only prediction model for neck pain that has been externally validated. However, the model has not been externally validated in settings other than general practice and physiotherapy in England and the Netherlands. This may be important as patient populations and predictors may differ across countries and settings.[13,14]

The aims of this study were therefore to (1) externally validate the prediction model developed by Schellingerhout (hereafter "the original model") in patients with neck pain consulting Norwegian chiropractors, (2) recalibrate the original model in chiropractic neck patients, (3) potentially update the original model by adding new predictors, and (4) internally validate the updated model.

MATERIALS AND METHODS

This study was part of a 1-year observational study following patients with neck pain consulting chiropractic practice. Decisions regarding treatment were at the discretion of the individual chiropractor, irrespective of study participation.

Setting

Altogether 71 members of the Norwegian Chiropractic Association agreed to recruit patients. They were located across all parts of Norway, reflecting urban and rural areas. Patients can be either referred or self-referred to chiropractic treatment and qualify for partial refund from the Norwegian health care system.[15]

Participants

Patients aged 18 to 70 years, presenting with neck pain as a primary or secondary complaint with or without arm pain were invited to participate. They were eligible for inclusion regardless of pain duration and if they had started treatment or not. Participants should be able to read and write Norwegian, and to respond SMS messages on a mobile phone (not used in this study). Exclusion criteria were suspicion of serious pathology or fracture as cause of neck pain. Chiropractors were instructed to invite all consecutive patients presenting with neck pain. The inclusion period was September 2015 to May 2016.

Procedure

All patients received oral and written information about the study from the chiropractor. All participants signed a written informed consent. The participants received questionnaires on paper or electronically. Paper questionnaires were given at recruitment by the chiropractor, and returned by the participant in a pre-paid envelope to the researchers. Participants choosing electronic questionnaires received an e-mail within 2 days with a link to the baseline questionnaire. Follow-up questionnaires were sent after 12 weeks. Those not responding within 7 days had one written reminder followed by a phone call 2 weeks later.

Patient Reported Baseline Variables

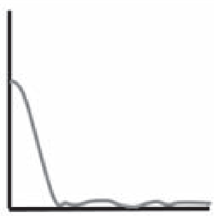
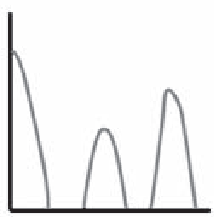
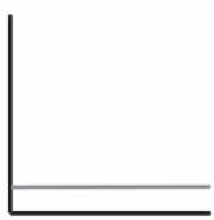
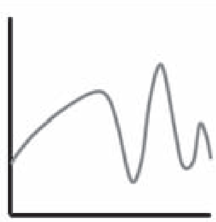
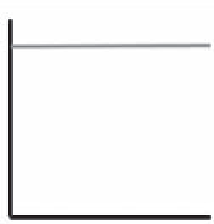
The baseline questionnaires included all nine predictors from the original model of Schellingerhout *et al*[12] (Table 1). In addition, demographic variables and other potential predictors used to update the model were selected based on the literature,[10,23–25] recommendations on prediction model development[26,27] and clinical experience. Hence, we included sum scores from five questionnaires.[16–21] Two of the additional variables were visual trajectory patterns describing the course of neck pain in the past year and expected course of neck pain in the forthcoming year. Five trajectories were made based on the literature of trajectory patterns[25,28] (Figure 1). All potential predictors were divided into five domains to be used when updating the model because many of the included potential predictors may carry similar or overlapping information (Table 1).

Consultation-Type

Consultation-type describes when in the course of treatment participants were recruited ("First-time consultation" = recruited at the first visit for a new episode of neck pain, "Follow-up consultation" = recruited during a clinical course of treatment, "Maintenance consultation" = patients visiting the chiropractor regularly at pre-planned time points[29]).

Outcome Measure

The outcome measure was self-reported global perceived effect (GPE) at 12 weeks.[30] GPE was measured on a 7-point Likert-scale (0 = "completely recovered" to 6 = "worse than ever"). Scores of 2–6 were coded as "persistent complaints" and used as outcome for the analysis as in the original model.[12]

1)		No neck pain or just a single episode of neck pain
2)		Few episodes of neck pain separated by pain free periods
3)		Mild neck pain most of the time
5)		Neck pain of varying intensity but never completely pain free
6)		Severe neck pain most of the time
7)	-	None of the above illustrated
8)	-	Do not know

Question 1 Previous course of neck pain: "Please tick off the description above that you think best represents how your neck pain has been the previous 12 months"
Question 2 Expected course of neck pain: "Please tick off the description above that you believe will represent how your neck pain will be for the following 12 months"

Figure 1. Self-reported visual trajectories of neck pain.

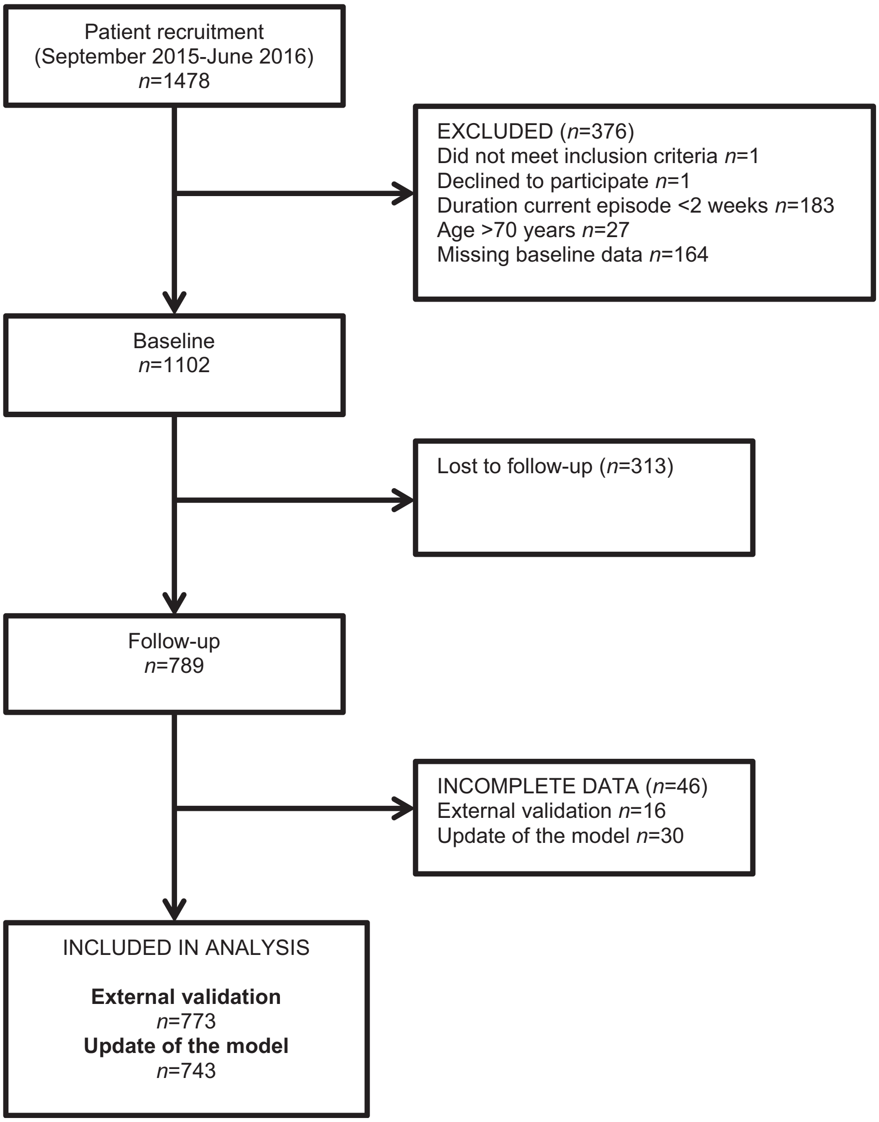


Figure 2. Flow chart of study participation.

The Data Sets and Approach for Validation and Update

First, the full study sample (n = 773) was used for external validation of the original model. Second, the full study sample was used for a recalibration of the original model. Third, the original model was updated using a randomly created split-sample from the full study sample stratified by number of recruited patients per clinic to achieve equal representation of clinics as in the full study sample (Development sample, n = 436). Fourth, the updated model was tested in the rest of the full study sample (Validation sample, n = 307), (Figures 2 and 3).

Descriptive analyses are presented as frequencies (%) or mean values with standard deviation (SD). Univariate logistic regression analysis was used to estimate relationships between the outcome and the variables Consultation-type and Duration current episode as well as their potential moderation of the relationship between original predictors and outcome. Only complete-cases where analyzed, thus we excluded 46 individuals having one or more single items missing of the predictors. All analyses were carried out using Stata version 15 (StataCorp., TX).

Stage 1. Independent External Validation of the Original Model

The original model[12] was applied in our full study sample (n = 773) using fixed coefficients, that is, by transporting the coefficients from the original model to the validation setting (Table 2). The validity of the model was evaluated in terms of calibration and discrimination.[20,31,32] Calibration was assessed graphically by the agreement between predicted and observed outcomes. The Hosmer-Lemeshow goodness-of-fit test is testing the null-hypothesis that observed and predicted outcomes do not differ.[33] Discrimination was assessed by area under the receiver-operating characteristic curve (AUC).

Stage 2. Recalibration of the Original Model

Recalibration of the original model coefficients (intercept and slope) was performed in the full study sample. The regression coefficients of the recalibrated model were evaluated as in Stage 1.

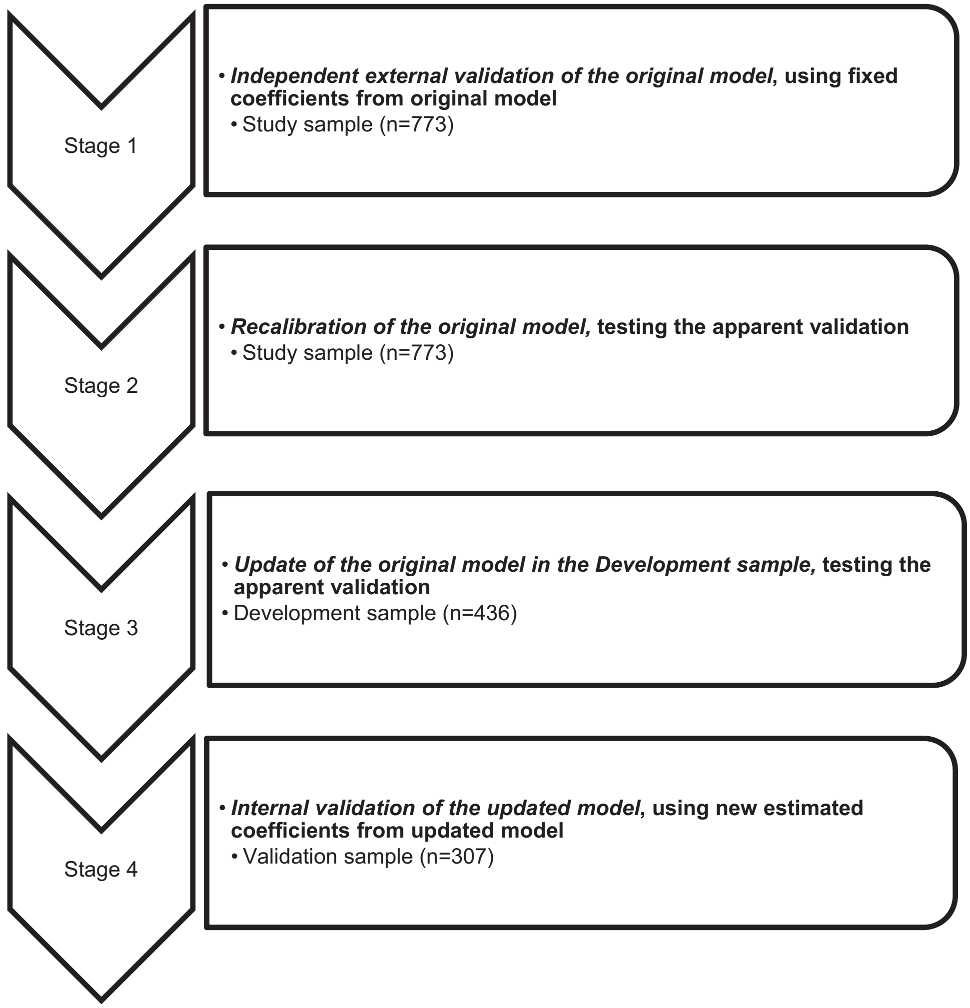


Figure 3. Overview of analyses stages.

Stage 3. Update of the Original Model in the Development Sample

The original model was updated using the development sample (n = 436). The update included recalibration of the coefficients of the predictors and removing/adding new predictors by a non-automated criterion-based procedure. First, original predictors were removed if the model fit was not significantly impaired (tested by the Likelihood Ratio test [LR] and Akaike's information criterion [AIC]). Starting with the interaction terms and followed by the individual predictors we removed those with an OR closest to one and with the largest *P*-values as long as the AIC value or LR were not negatively affected. The updated model with the lowest AIC value and an unchanging LR was chosen. Within the five domains described in Table 1, all potential predictors were included in logistic regression models using GPE as outcome and removed one by one based on AIC and LR.

The best performing predictor(s) from each domain were then included in the updated model. Finally, predictors were removed from the updated model using the same non-automated procedure as for updating the original model. The performance of the updated model was evaluated in terms of calibration and discrimination. The non-automatic approach was chosen to avoid unstable variable selection from stepwise methods.[34] A sufficiently large sample size provided more stable estimates of model performance.[35–37]

Stage 4. Internal Validation of the Updated Model

Reproducibility of the updated model obtained from Stage 3 was tested using the Validation sample (n = 307), (internal validation). The updated model was evaluated in terms of calibration and discrimination as in Stage 1.

The study was approved by Regional committees for medical and health research ethics (2015/89).

RESULTS

Patient Characteristics

Altogether1478 patients were recruited of which 1102 met the inclusion criteria and were included in this study. Dropouts were 28% (n = 313) for outcome measures at 12-week follow-up. Hence, the full study sample consisted of 773 patients (Figure 2). The baseline characteristics of our full study sample, dropouts and the original study are presented in Table 3. Participants included in the analyses and those lost to follow-up had similar sociodemographic and neck

TABLE 2. Parameter Estimates for the Original Sample (see [12]) and the Re-Calibrated Original Model using Logistic Regression Analysis in Stage 2

Development Sample, n = 436				
	Updated Model			
Predictor	**β-Coefficients**	**SE**	**OR**	**95% CI**
Constant	−1.061	1.05		
Radiating pain to shoulder and/or elbow	−0.689	0.282	0.50	0.29–0.87
Education level				
Low	Reference (Ref.)			
Medium	−1.302	0.671	0.27	0.07–1.01
High	−1.08	0.665	0.34	0.09–1.25
Physical activity	−1.583	0.604	0.21	0.06–0.67
Consultation-type				
First-time consultation	Ref.			
Follow-up consultation	0.553	0.414	1.74	0.77–3.91
Maintenance consultation	0.773	0.392	2.17	1.00–4.67
Patient previous trajectory pattern				
Single episode	Ref.			
Episodic pain	1.546	0.679	4.69	1.24–17.75
Mild pain/recovering	1.742	0.736	5.71	1.35–24.15
Fluctuating pain	1.821	0.673	6.18	1.65–23.08
Moderate/severe pain	1.580	0.913	4.86	0.81–29.08
Neither/ Unsure	1.512	0.954	4.54	0.7–29.43
Number of pain sites	0.017	0.099	1.02	0.84–1.23
Patient expected trajectory pattern				
Single episode	Ref.			
Episodic pain	0.578	0.36	1.78	0.88–3.60
Mild pain/recovering	0.787	0.447	2.2	0.92–5.28
Fluctuating pain	1.144	0.415	3.14	1.39–7.08
Moderate/severe pain	0.084	1.420	1.09	0.07–17.6
Neither/Unsure	0.431	0.425	1.54	0.67–3.54
Physical activity number of pain sites	0.219	0.112	1.24	1.00–1.55

pain related variables. Participants of the original study had a higher fraction of male participants (39% *vs.* 25%), a lower education level, a smaller fraction reporting previous neck complaints (64% *vs.* 87%) and a lower prevalence of low back pain (21% and 55%). Persistent complaints at 12 weeks in our study was 47% compared with 43% after 6 months in the original study.

Stage 1. Independent External Validation of the Original Model

The original model showed poor discriminative ability: AUC (95% CI) = 0.55 (0.51–0.59) (Figure 4) and the calibration plot showed a poor fit of the model to the data (Figure 5) (Hosmer-Lemeshow test $P < 0.001$). Key results of validation and updating during all stages are shown in Table 4, Figures 4 and 5.

Stage 2. Recalibration of the Original Model

The performance improved after recalibration of the original model; see Tables 2 and 4 for details. The AUC increased (Figure 4) and the calibration plot showed a clear improvement in precision (Figure 5). Pain intensity and radiating pain showed a stronger association with outcome in our sample as compared with the original study, whereas low back pain was less predictive. None of the interaction terms from the original model were significantly associated with outcome (Table 2).

Stage 3. Update of the Original Model in the Development Sample

The updated model included seven predictors and one interaction term; Table 1 shows excluded and included predictors of the updated model and Table 5 show parameters of the updated model. The updated model included radiating pain to shoulder and/or elbow, education level, physical activity, consultation-type, expected course of neck pain, previous course of neck pain, number of pain sites, and the interaction term physical activity##number of pain sites. The only original predictor remaining in the updated model was radiating pain. The performance of the updated model

TABLE 3. Baseline Characteristics of the Original Sample (see [12]) and the Present Chiropractic Study Sample (n = 773)

	Original Sample	Present Study Sample	Present Study Dropouts	Development Sample	Validation Sample
	n = 468	n = 773	n = 313	n = 436	n = 307
Domain 1: Sociodemographic and physical factors					
Gender, (male, %)	182 (39)	193 (25)	80 (26)	113 (26)	73 (24)
Age, mean, years (SD*)	45.4 (11.8)	45.1 (12)	41.6 (12.4)	44.7 (11.6)	44.9 (12.2)
Range	18–70	18–70	18–70		
Smoking, (yes, %)		74 (9.6)	46 (15)	40 (9.2)	30 (9.8)
Marital status, (n, %)					
Married		592 (77)	233 (74)	347 (80)	222 (72)
Divorced		55 (7)	19 (6)	26 (6)	27 (9)
Widow (er)		8 (1)	2 (1)	3 (0.7)	5 (1.6)
Single		118 (15)	59 (19)	60 (14)	53 (17)
Education level†, (n, %)					
High	135 (29)	442 (57)	146 (47)	265 (61)	159 (52)
Medium	195 (42)	298 (39)	150 (48)	155 (36)	134 (44)
Low	122 (26)	33 (4)	15 (5)	16 (4)	14 (5)
Employment status, (employed, %)	334 (71)	635 (82)	252 (81)	359 (82)	252 (82)
Biking, (yes, %)		72 (9)	28 (9)	42 (10)	25 (8)
Physical activity, (≥1 per week, %)		538 (70)	215 (69)	305 (70)	213 (69)
Domain 2: Clinician reported					
Consultation-type, (n, %)					
First-time consultation		101 (13)	58 (19)	48 (11)	50 (16)
Follow-up consultation		230 (29)	93 (30)	129 (30)	95 (31)
Maintenance consultation		437 (55)	152 (49)	259 (59)	162 (53)
Domain 3: Neck pain history					
Previous neck complaints, (yes, %)	301 (64)	672 (87)	259 (83)	381 (87)	265 (86)
Traumatic cause neck complaint, (trauma, %)	63 (14)	112 (14)	32 (10)	66 (15)	43 (14)
Radiating pain to shoulder and/or elbow, (yes, %)	296 (63)	590 (76)	238 (76)	319 (73)	250 (81)
Pain intensity (NRS, 0–10), median (IQR‡)	5.7 (2.1)	4 (3–6)	5 (3–6)	4 (3–6)	4 (2–6)
Pain intensity = 0 at baseline, (n, %)		33 (4)	9 (3)	26 (6)	6 (2)
Duration current episode, (n, %)					
Less than <1 month	58 (13)	76 (10)	35 (11)	42 (10)	30 (10)
1–3 months	225 (48)	129 (16)	57 (18)	65 (15)	60 (20)
>3 months	160 (34)	567 (73)	214 (68)	329 (75)	217 (71)
NDI (0–50-scale), (mean, SD)	14.5 (6.7)	11.8 (6.4)	12.8 (6.9)	11.9 (6.5)	11.7 (5.9)
Onset of neck pain, (n, %)					
Acute		127 (16)	34 (11)	69 (16)	53 (17)
Over time		581 (75)	248 (79)	327 (75)	234 (76)
Do not know		64 (8)	31 (10)	40 (9)	20 (7)
Use of painkillers, (yes, %)		451 (58)	180 (58)	252 (58)	184 (60)
Patient previous trajectory pattern (n, %)					
Single episode		54 (7)	17 (5)	35 (8)	18 (6)
Episodic pain		237 (30)	85 (27)	131 (30)	95 (31)
Mild pain/recovering		75 (10)	31 (10)	40 (9)	31 (10)
Fluctuating pain		367 (47)	147 (47)	205 (47)	149 (49)
Moderate/severe pain		20 (3)	13 (4)	15 (3)	4 (1)
Neither/Unsure		16 (2)	12 (4)	10 (2)	10 (3)
Sick-listing for present neck pain, (yes, %)		41 (5)	22 (7)	23 (5)	17 (6)
Insurance claim of present neck pain, (yes, %)		43 (5)	23 (7)	16 (4)	24 (8)

TABLE 3 (*Continued*)

	Original Sample	Present Study Sample	Present Study Dropouts	Development Sample	Validation Sample
	n = 468	n = 773	n = 313	n = 436	n = 307
Domain 4: General health/Comorbidity					
Low back pain (NPQ past 7 days), (yes, %)	96 (21)	423 (55)	178 (57)	250 (57)	158 (51)
Headache (NPQ past 7 days), (yes, %)	317 (68)	573 (74)	222 (71)	329 (75)	224 (73)
Number of pain sites, (NPQ sum score, 0–10), mean (SD)		4.8 (2.1)	4.7 (2.3)		
Health status, (Eq5d 100 mm VAS), mean (SD)	69.9 (17.3)	70.2 (19.4)	68.2 (20.6)	69.2 (20.2)	71.5 (17.9)
Eq5d-index, mean (SD)		0.85 (0.12)	0.84 (0.13)	0.85 (0.13)	0.86 (0.10)
Domain 5: Psychological					
TAMPA-scale, mean (SD)[§]	33.8 (7.1)	-	-	-	-
Kinesiophobia, (single question), mean (SD)		2.3 (2.7)	2.7 (2.9)	2.4 (2.8)	2.1 (2.5)
HSCL-score, mean (SD)		1.6 (0.5)	1.7 (0.5)	1.7 (0.5)	1.6 (0.5)
Örebro score, mean (SD)		41.1 (14.4)	42.4 (16.2)	41.6 (14.4)	40.7 (14)
Patient expected trajectory pattern (n, %)					
Single episode		135 (17)	51 (16)	80 (18)	51 (17)
Episodic pain		274 (35)	100 (32)	142 (33)	121 (39)
Mild pain/recovering		97 (13)	37 (12)	56 (13)	37 (12)
Fluctuating pain		164 (21)	80 (25)	95 (22)	64 (21)
Moderate/severe pain		3 (0.4)	6 (2)	3 (1)	5 (2)
Neither/ Unsure		99 (13)	38 (12)	60 (14)	29 (9)

*SD standard deviation.

†High = University College or university over 4 years, medium = University College or university under 4 years or upper secondary school/high school, low = vocational school or lower secondary school.

‡IQR interquartile range.

§Did not obtain Tampa in one of the RCT's.

NRS: the 11-point numerical rating scale,[20] NDI: the Neck Disability Index measuring disability,[21] NPQ: Nordic pain questionnaire measuring musculoskeletal symptom prevalence,[19] Eq5d: Health-related quality of life,[22] TAMPA scale, HSCL-10: measuring anxiety and depression,[16] Örebro Screening Questionnaire: predicting long-term disability and failure to return to work.[17,18]

improved compared with the original and to the recalibrated model with respect to both discriminative ability and calibration (Figures 4 and 5).

Stage 4. Internal Validation of the Updated Model

The updated model had reasonable discriminative ability: AUC (95% CI) = 0.65 (0.58–0.71) in the validation sample (n = 307) see Figure 4. Calibration plot predicted best those at low-risk of persistent pain (Figure 5) but the Hosmer-Lemeshow test was significant ($P < 0.01$).

DISCUSSION

Our study showed poor external validity of the original model in Norwegian chiropractic patients with neck pain. Performance was improved through recalibration of the model. However, the model was still not able to predict GPE well in this setting. During the update, all predictors from the original model apart from radiating pain were excluded and replaced with new ones resulting in a model so different from the original one that new external validation is needed.

The original model had previously been externally validated with a reasonable discrimination and an acceptable calibration.[12] Our study was a large prospective cohort study planned for the external validation with all original predictors and outcome collected and categorized similar to the original study. Some methodological differences might be of importance. Outcome was measured after 12 weeks rather than 6 months in the original study. Still, this disparity did not give any major difference in the prevalence of the outcome, 47% *versus* 43%. As most substantial improvements occur within 6–12 weeks after care seeking,[38–40] it is expected that a model, being predictive of 6-months outcome, would also predict 3-months outcomes. We only included complete cases instead of imputing missing data.[27,41] This may have introduced bias, but with a small number of missing items (<3%) we believe this to be of minimal importance.

The lack of external validity in our study could be due to differences between the two samples, indicated by different distributions of predictors between the populations. The original model was developed using patients participating in a RCT, but applied in our study sample with looser inclusion

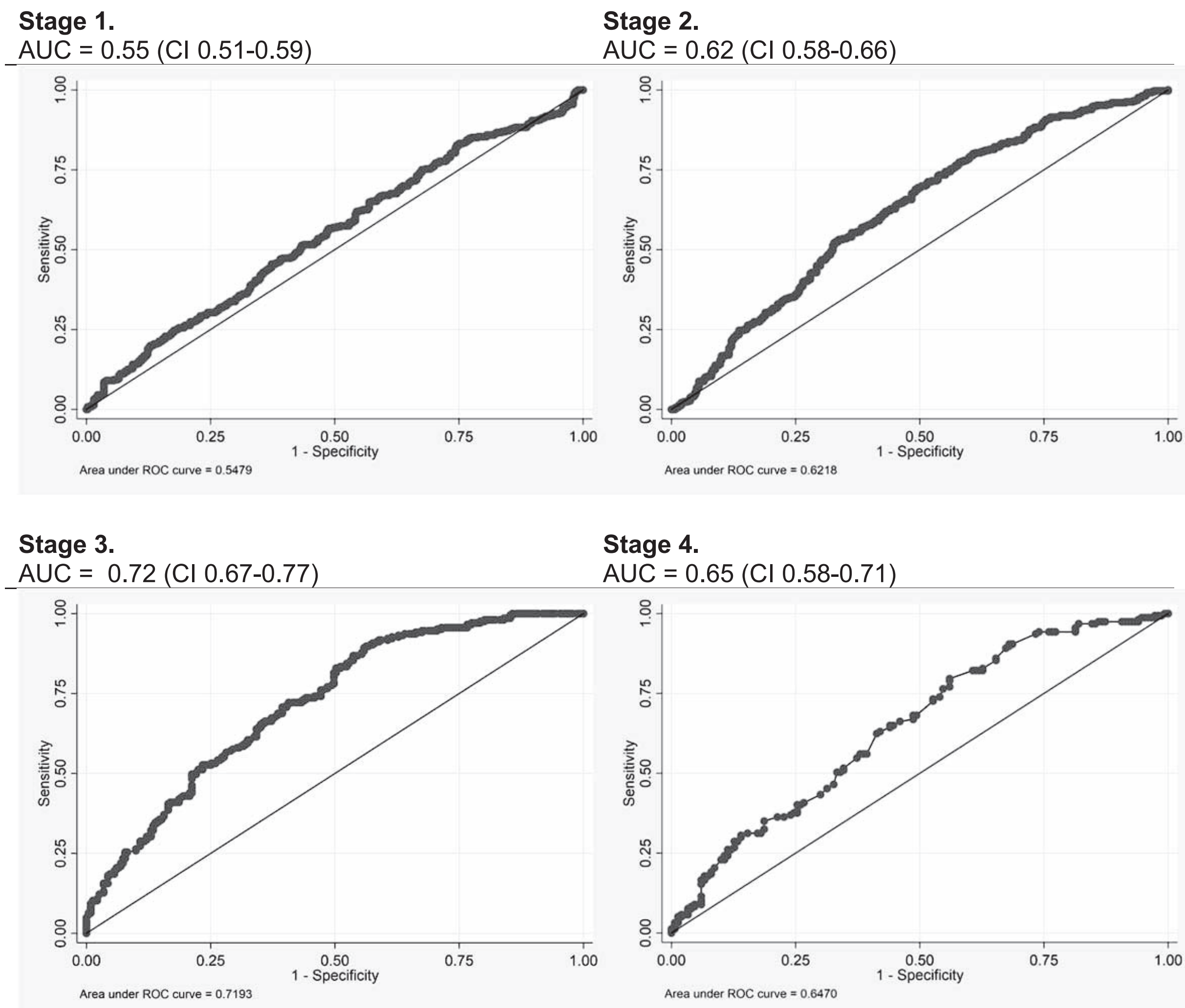


Figure 4. Area under the Receiver operator characteristic curve (AUC) showing the ability of the different prediction models to correctly classify those with and without persistent pain in Stage 1–4. AUC = 0.5 indicates no discrimination whereas AUC = 1.0 indicates perfect discrimination. The solid reference line refers to no discrimination.

criteria. Different populations might also be expected due to the settings; the original model developed in general practice and physiotherapy and our validation in chiropractic setting. The two cohorts differed on sex, education, previous complaints, and presence of low back pain, which are all factors known to be associated with persistent neck pain.[10] Furthermore, less restricted criteria resulted in twice as many participants with pain duration more than 3 months compared with the original model. This difference may affect prediction since a longer duration of the symptoms at baseline is related to poor outcome.[42,43] However, the associations between predictors and outcome were not substantially different between groups differing in duration and we believe the longer pain duration explains only little of the poor performance. It is also possible that there were differences in psychosocial factors between the samples, although the lack of such data in the original study precludes further discussion of this.

The present model update included three novel predictors that seemed to be stronger than some of the predictors commonly used. Previous and Expected visual trajectory patterns both showed strong association with outcome. This may indicate that patients' recall of their overall symptom history may carry more prognostic information than traditional measures of previous episodes and episode duration.[43] This fits well with emerging evidence on quite stable long-term trajectories of spinal pain.[44,45] Inclusion of the predictor Expected course of neck pain in the updated model should possibly be interpreted as a proxy measure of patients' expectations and possible psychological distress.[25] It is somewhat surprising that this variable was a stronger predictor than the traditional variables reflecting psychosocial factors. These findings need further investigation regarding prognostic information of the trajectory patterns.

It is also noteworthy that the variable Consultation-type had prognostic effect. Patients receiving maintenance care have poorer outcome after 3 months compared with those included with a new episode. This is not surprising, but suggests that this variable includes information of prognostic value that complements the other variables. Further

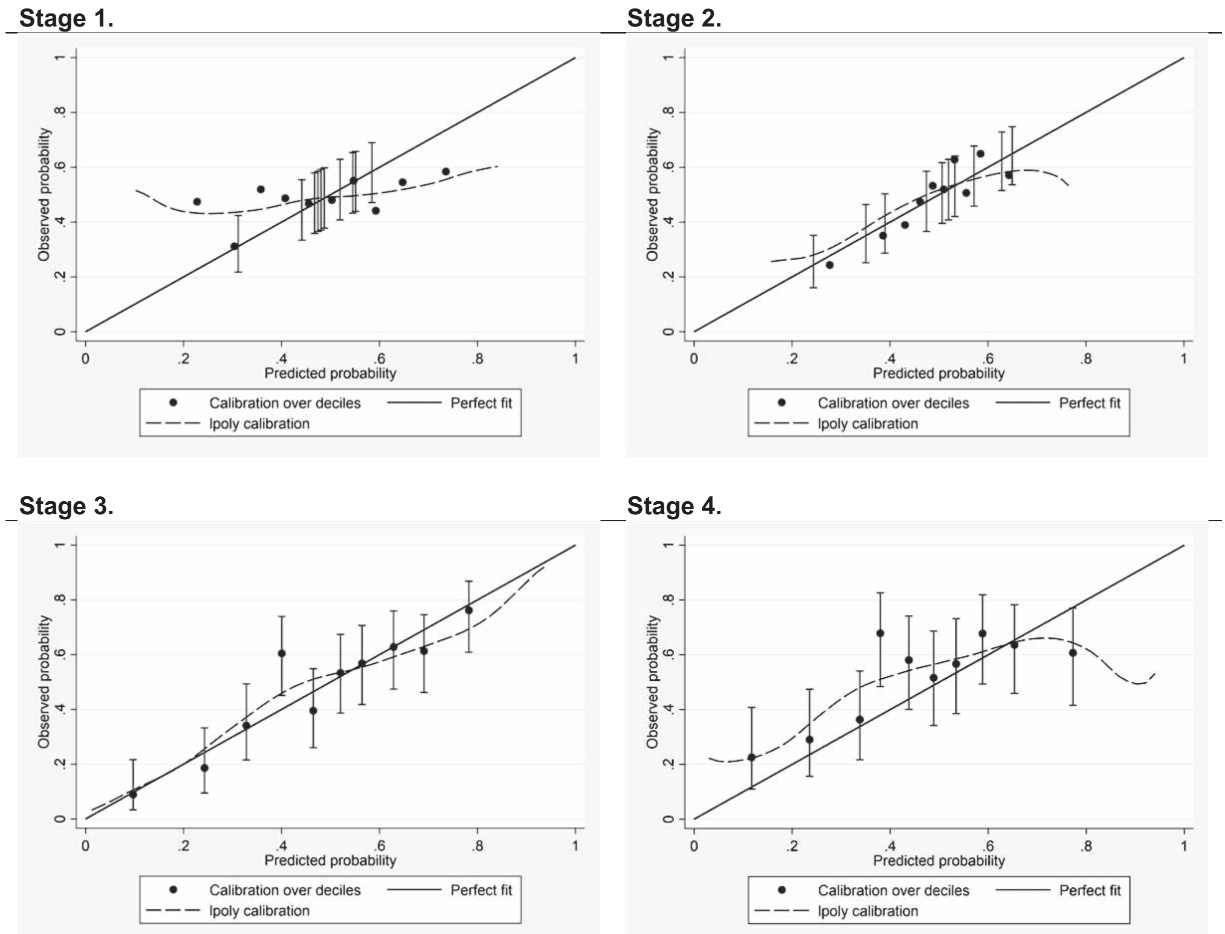


Figure 5. Calibration plots with distribution of observed frequencies and predicted probabilities at Stage 1–4. The calibration curve is estimated by local regression (lpoly). The solid reference line indicates perfect fit of the prediction model.

investigation is needed to understand how this variable affects outcome and whether similar effects can be found for patients in other settings.

The inclusion of the novel predictors provided a model that performed best in identifying people with a low probability of persistent pain. This will be of value in reducing overtreatment of patients with a good prognosis. The performance of the model for identifying persons with high probability of persistent pain was less informative. The significant Hosmer-Lemeshow test should be interpreted with caution because a large sample size increases the probability of a statistically significant lack of fit. The updated model should thus be further evaluated and not rejected solely based on this test.

TABLE 4. Summary of Model Performance Measures off the Original Sample (see [12]) and all Stages

	Original Sample	Stage 1	Stage 2	Stage 3	Stage 4
	Original Model	Original Model	Original Model	Updated Model	Updated Model
	Original Sample	Full Study Sample	Full Study Sample	Development Sample	Validation Sample
	n = 468	n = 773	n = 773	n = 436	n = 307
Discrimination, AUC (95% CI)	0.66 (0.61–0.71)*	0.55 (0.51–0.59)	0.62 (0.58–0.66)	0.72 (0.67–0.77)	0.65 (0.58–0.71)
Calibration, Hosmer-Lemeshow	$P = 0.61^{\dagger}$	$P < 0.001$	$P = 0.36$	$P = 0.23$	$P = 0.003$
Number of events/proportion with persistent pain	43%	49%	49%	47%	51%

*Value from equation.

†Values from score chart.

TABLE 5. Parameter Estimates for the Updated Model using Logistic Regression Analysis in Stage 3

	Original Sample, n = 468				Present Study Sample, n = 773			
	Original Model				Recalibrated Original Model			
Predictor	β-Coefficients	SE	OR	84.3% CI*	β-Coefficients	SE	OR	95% CI
Constant	−1.704	0.838			−1.136	0.725		
Age	0.029	0.015	1.03	1.01–1.05	0.002	0.006	1.00	0.99–1.01
Employment status	0.372	0.244	1.45	1.03–2.05	0.297	0.405	1.34	0.61–2.98
Previous neck complaint	0.515	0.207	1.67	1.25–2.24	0.124	0.435	1.13	0.48–2.66
Traumatic cause neck complaint	0.234	0.149	1.26	1.02–1.56	0.223	0.211	1.25	0.83–1.89
Radiating pain to shoulder and/or elbow	−0.564	0.212	0.57	0.42–0.77	0.115	0.377	1.12	0.54–2.35
Pain intensity	−0.042	0.058	0.96	0.88–1.04	0.216	0.082	1.24	1.06–1.46
Low back pain	0.829	0.415	2.29	1.27–4.12	0.251	0.151	1.29	0.96–1.73
Headache	0.198	0.498	1.22	0.60–2.46	0.985	0.693	2.68	0.69–10.43
Health status, (Eq5d)	−0.005	0.003	1.00	0.99–1.00	−0.008	0.004	0.99	0.98–1
Headache pain intensity	0.116	0.073	1.12	1.01–1.24	−0.182	0.092	0.83	0.7–1
Headache previous neck complaint	−0.376	0.252	0.69	0.48–0.98	0.207	0.505	1.23	0.46–3.31
Headache radiating pain to shoulder and/or elbow	0.392	0.253	1.48	1.03–2.12	−0.135	0.431	0.87	0.38–2.03
Headache employment status	−0.815	0.276	0.44	0.30–0.65	−0.189	0.458	0.83	0.34–2.03

Schellingerhout et al[12] used a 84.3% confidence interval as this corresponds with the criterion ($P < 0.157$) during the selection procedure.

Our results raise the question whether it is realistic that prediction models from one setting have the potential for use in other settings. In acute low back pain, promising results imply that the same model in Australian general practice and Danish chiropractic can identify people with a good outcome of care.[46] This does not preclude, however, that even better prognosis could be obtained in these settings with another model. One might even hypothesise that models for this kind of health service with a wide set of treatment options, should be individualized to the therapist.

This study is one of few to independently externally validate a prediction model for neck pain.[11] We did not find that the original model was predictive in this sample of patients managed by chiropractors. The between-population-heterogeneity might be a limitation when transferring prediction models to different settings. An attempt to update the model resulted in a new prediction model that was able to predict patients with a favorable outcome. It is, however still pre-mature to be used in clinical decision-guidance and would need further evaluation and perhaps updating before implementation is considered.

➢ Key Points

- A previously validated prediction model (called original model) was externally validated in 773 chiropractic patients with neck pain.
- The performance of the external validation of the original model was poor.
- An update of the original model included the predictors: radiating pain to shoulder and/or elbow, education level, physical activity, consultation-type, expected course of neck pain, previous course of neck pain, number of pain sites, and the interaction term physical activity##number of pain sites.
- The updated prediction model performed best in identifying people with a low probability of persistent pain but need further evaluation.
- The updated prediction model included patient-reported visual trajectories of neck pain pattern that should be investigated further regarding prognostic information.

Acknowledgments

The authors thank Knut Waagan for his help during statistical analyses. They also thank the participating chiropractors and patients.

References

1. Hogg-Johnson S, et al. The burden and determinants of neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)* 2008;33 (4 suppl):S39–51.
2. Fejer R, Kyvik KO, Hartvigsen J. The prevalence of neck pain in the world population: a systematic critical review of the literature. *Eur Spine J* 2006;15:834–48.

3. Hoy DG, et al. The epidemiology of neck pain. *Best Pract Res Clin Rheumatol* 2010;24:783–92.
4. Vos T, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380:2163–96.
5. Natvig B, et al. Musculoskeletal symptoms in a local community. *Eur J Gen Pract* 1995;1:25–8.
6. Graham N, et al. An ICON overview on physical modalities for neck pain and associated disorders. *Open Orthop J* 2013;7: 440–60.
7. DALYs GBD, Collaborators H. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017;390:1260–344.
8. Lærum E, Brage S, Ihlebæk C, Johnsen K, Natvig B, & Aas E. A musculoskeletal accounting: Prevalence and costs related to injuries, disease and disorders of the musculoskeletal system: MST-report. 1/2013. MST-report. ISBN 978-82-303-2311-3 ed: The Norwegian council for musculoskeletal health by/ FORMI, Division of surgery and neuroscience, Oslo university hospital - Ullevål. Postbox 4956 Nydalen, 0424 Oslo, Norway 2013:21–3.
9. Hurwitz EL, et al. Treatment of neck pain: noninvasive interventions: results of the bone and joint decade 2000-2010 task force on neck pain and Its associated disorders. *Spine (Phila Pa 1976)* 2008;33 (4 suppl):S123–52.
10. Carroll LJ, et al. Course and prognostic factors for neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)* 2008;33 (4 suppl):S75–82.
11. Kelly J, Ritchie C, Sterling M. Clinical prediction rules for prognosis and treatment prescription in neck pain: a systematic review. *Musculoskelet Sci Pract* 2017;27:155–64.
12. Schellingerhout JM, et al. Prognosis of patients with nonspecific neck pain: development and external validation of a prediction rule for persistence of complaints. *Spine (Phila Pa 1976)* 2010;35: E827–35.
13. Hestbaek L, et al. Low back pain in primary care: a description of 1250 patients with low back pain in danish general and chiropractic practice. *Int J Family Med* 2014;2014:106102.
14. Morso L, et al. The predictive and external validity of the STarT Back Tool in Danish primary care. *Eur Spine J* 2013;22:1859–67.
15. Legislation regarding changes in treatment reimbursement for chiropractors and manual therapists, the ministry of health and care services: proposition no 28 (Ot. prp. nr. 28), 2011.]. Available at: http://www.stortinget.no/no/Saker-og-publikasjoner/Vedtak/Beslutninger/Odelstinget/2005-2006/beso-200506-022/-a1. Accessed June 29, 2019.
16. Derogatis LR, et al. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 1974;19:1–15.
17. Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. *Clin J Pain* 2003;19:80–6.
18. Grotle M, Vollestad NK, Brox JI. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. *Clin J Pain* 2006;22:458–67.
19. Kuorinka I, et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon* 1987;18: 233–7.
20. Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. *Spine (Phila Pa 1976)* 2000;25:3140–51.
21. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991;14:409–15.
22. EuroQol G. EuroQol–a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
23. Kongsted A, et al. What have we learned from ten years of trajectory research in low back pain? *BMC Musculoskelet Disord* 2016;17:220.
24. Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. *Am J Epidemiol* 2006;163:754–61.
25. Dunn KM, Campbell P, Jordan KP. Validity of the visual trajectories questionnaire for pain. *J Pain* 2017;18:1451–8.
26. Steyerberg EW. Clinical prediction models. A practical approach to development, validtion, and updatinged. Springer Science+° Business Media, LLC, 233 Spring Street, New York, NY 10013, USA: Springer, 2009.
27. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validtion, and Updating. 2009: Springer.
28. Kongsted A, Hestbaek L, Kent P. How can latent trajectories of back pain be translated into defined subgroups?. *BMC Musculoskelet Disord* 2017;18:285.
29. Myburgh C, et al. The Nordic maintenance care program: what is maintenance care? Interview based survey of Danish chiropractors. *Chiropr Man Therap* 2013;21:27.
30. Kamper SJ, et al. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol* 2010;63:760.e1–6.e1.
31. Steyerberg EW, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
32. Moons KM, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): explanation and elaboration. *Ann Int Med* 2015;162:W1–73.
33. Hosmer DW Jr, Lemeshow S. *Applied Logistic Regression*, Second edition ed. New York: Wiley-Interscience Publication; 2000.
34. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996;49:907–16.
35. Kent P, et al. Four hundred or more participants needed for stable contingency table estimates of clinical prediction rule performance. *J Clin Epidemiol* 2017;82:137–48.
36. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.
37. Vergouwe Y, et al. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–83.
38. Leaver AM, et al. People seeking treatment for a new episode of neck pain typically have rapid improvement in symptoms: an observational study. *J Physiother* 2013;59:31–7.
39. Walton DM, et al. Exploring the clinical course of neck pain in physical therapy: a longitudinal study. *Arch Phys Med Rehabil* 2014;95:303–8.
40. Vasseljen O, et al. Natural course of acute neck and low back pain in the general population: the HUNT study. *Pain* 2013;154: 1237–44.
41. Janssen KJ, et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol* 2010;63:721–7.
42. Bot SD, et al. Predictors of outcome in neck and shoulder symptoms: a cohort study in general practice. *Spine (Phila Pa 1976)* 2005;30:E459–70.
43. Bruls VE, Bastiaenen CH, de Bie RA. Prognostic factors of complaints of arm, neck, and/or shoulder: a systematic review of prospective cohort studies. *Pain* 2015;156:765–88.
44. Dunn KM, Campbell P, Jordan KP. Long-term trajectories of back pain: cohort study with 7-year follow-up. *BMJ Open* 2013;3: e003838.
45. Lemeunier N, Leboeuf-Yde C, Gagey O. The natural course of low back pain: a systematic critical literature review. *Chiropr Man Therap* 2012;20:33.
46. da Silva T, Macaskill P, Kongsted A, et al. Predicting pain recovery in patients with acute low back pain: Updating and validation of a clinical prediction model. *Eur J Pain* 2019;23:341–53.