

Predicting Short-Term Response and Non-Response to Neck Strengthening Exercise for Chronic Neck Pain

Jennifer Lyn Keating
Peter Kent
Megan Davidson
Ross Duke
Lachlan McKinnon
Robert De Nardis

ABSTRACT. Summary of Background Data. People with chronic neck pain respond variably to exercise therapy. The likely success of exercise for chronic neck pain cannot be predicted.

Study Design. This prospective study tested the relationship between patient characteristics assessed prior to intervention and response to an exercise program for chronic neck pain.

Objectives. To investigate whether responders differed from non-responders with respect to presenting characteristics.

Method. Data were collected on 336 consecutive patients receiving physiotherapy for chronic neck pain. Subject age, sex, duration of symptoms, compensation status, Neck Disability Index (NDI) item and total scores, cervical spine range of movement, and isometric strength were

Jennifer Lyn Keating, PT, PhD, Peter Kent, DC, PT, GradDipManipTher, Megan Davidson, PT, PhD, Ross Duke, PT, MPhysio, Lachlan McKinnon, PT, MManipTher, are all affiliated with the School of Physiotherapy, La Trobe University.

Robert De Nardis, PT, BAppSc (Phy), is affiliated with the Melbourne Whiplash Centre.

Address correspondence to: Jennifer Lyn Keating, School of Physiotherapy, La Trobe University, Bundoora 3086, Australia (E-mail: j.keating@latrobe.edu.au).

The authors are grateful to the Faculty of Health Science, La Trobe University, for providing funds to support this project.

assessed prior to treatment. Response to therapy was defined as a change of 14/100 or more NDI points. Subjects received a progressive, concentric/eccentric, strengthening program using computerized equipment designed for cervical spine assessment and rehabilitation. Logistic regression formulae for predicting outcome at the completion of the program were developed on one sample ($n = 122$) data and tested on another ($n = 214$).

Results. In the first sample, NDI scores and item responses predicted response or non-response with approximately 70% accuracy. When tested on the second sample, a prediction model using NDI item scores predicted responders (positive predictive value 56%) and non-responders (negative predictive value 74%) with considerable loss of responder prediction accuracy. Participants with low initial NDI scores (higher functional ability) appeared to benefit less from the exercise program.

Conclusions. Functionally able patients may benefit less from exercise for chronic neck pain than those with greater functional limitations. [Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <<http://www.HaworthPress.com>> © 2005 by The Haworth Press, Inc. All rights reserved.]

KEYWORDS. Chronic neck pain, exercise, predicting response

Chronic neck pain affects many people and may significantly disable as many as 4.6% (95% CI, 3.3-5.8) (1). The ability of treating clinicians to discriminate between effective and ineffective treatment is particularly important in chronic musculo-skeletal conditions. The burden of chronic disease is not well served by ineffective treatment. Identifying those who are unlikely to respond to treatment should precede research into treatment effectiveness. If participants of randomized controlled trials include those with a low probability of response to the intervention under investigation, the potential utility of the treatment for those who might respond could be underestimated or overlooked. In addition, there are ethical and social obligations to provide patients with interventions that are both beneficial and cost-effective.

Although, anecdotally, clinicians consider that neck strengthening exercise for people with chronic neck pain may be worthwhile, scientific support for positive effects of such programs is limited. Verhagen et al. (2) conducted a Cochrane review of conservative treatments for whiplash and concluded that, although active interventions appear pref-

erable to passive interventions, the specific effects of exercise were unclear. In 2001, Hoving et al. (3) conducted a systematic review to assess concordance among reviews on conservative treatment of neck disorders. They found consensus that exercise for neck pain was of unknown benefit.

Participants in investigations of strengthening programs for chronic neck pain have not benefited uniformly. Some participants appear to respond well and others little or not at all. Stratford et al. (4) argued that change in neck disability index (NDI) scores of 14/100 points or more indicates change that is important to the patient, the therapist or both. Using the results of three randomized controlled trials (RCTs) of exercise for neck pain, we investigated how much change typically occurred in NDI scores, and the percentage of trial participants who changed their scores by at least 14 NDI points.

Bronfort et al. (5) conducted a RCT comparing neck exercise and spinal manipulation for people with 12 or more weeks of neck pain. Patients were given either manipulation with strengthening exercises, exercise delivered via a cervical spine extension and rotation machine, or manipulation alone. They found no significant differences between the effects of the three interventions on NDI scores. The following analysis was conducted using data for those who received exercise alone. After 5 weeks of the program, the mean change score was 9.6 NDI points. We estimated the correlation between initial and 5-week NDI scores to be 0.67 using our clinical data and calculated the standard deviation of change scores (6). Based on these figures, as many as 70% of subjects did not change by 14 or more NDI points. After 11 weeks of treatment, the mean change score was 14.3 NDI points, indicating that approximately 48% of participants may not have improved by an important amount.

Ylinen et al. (7) studied people with at least 6 months of constant or frequently-occurring neck pain. Participants were allocated to control, endurance, or strength training. After an institution-based training program, participants continued to perform home exercises for a year. At 12 months, the median improvement for control subjects who performed general aerobic and stretching exercises was 3 NDI points (95% CI 0-6). The median improvement was 8 NDI points (95% CI 6-11) for the endurance training group and 9 (95% CI 7-11) for the strength training group. Both exercise groups improved significantly more than the control group. The NDI change scores were not normally distributed so the percentage of participants who changed by 14 NDI points could not

be determined. It appears, however that at least 50% of the sample may not have experienced important change in response to exercise.

Recently, Korthals-de Bos et al. (8) reported a RCT comparing spinal mobilization, exercise, or general practitioner care for people with neck pain for at least two weeks. Seventy percent of participants had neck pain for less than 12 weeks. Differences between groups for changes in NDI scores at 12 months were not significant. Converting the data reported by Korthal-de Bos et al., to percentage NDI points, the mean change for the exercise group was 12.6 NDI points (SD 16) indicating that up to 53% (95% CI 43-63) of participants may not have achieved important functional changes. In addition, when directly surveyed by Korthal-de Bos et al., 47% of participants reported that they had not made important change. These data consistently indicate that there are some people with neck pain who do not appear to achieve functional improvements in response to neck strengthening programs.

The unpredictability in response to therapy may be due to the varying nature of injuries or to differences in the type or quality of treatment that is provided. However, differential diagnosis of the cause of chronic neck pain is not possible in the majority of cases. If neck pain is the result of a range of conditions, it is not surprising that some people might respond poorly to a particular intervention, while others do well. Health care providers and their clients would be advantaged if they could recognize those likely to benefit from an exercise regime before commencing that course of treatment. Currently, however, the presenting characteristics that indicate the potential for people with chronic neck pain to improve with an exercise program have not been identified.

This research aimed to examine the response to an individually tailored exercise program for chronic neck pain in a population of consecutive patients seeking treatment at a physiotherapy clinic. The specific research objectives were to identify the proportion of patients who changed by at least 14 NDI points, determine whether patients who responded and who did not respond to exercise differed with respect to presenting characteristics, and develop equations for predicting the likelihood of response to therapy for individual patients.

METHODS

The La Trobe University Human Ethics Committee approved the study design.

From 1998 to 2002, as part of their usual management of patients seeking treatment for chronic neck pain, physiotherapists at an Australian clinic collected pre-intervention data on characteristics of people presenting for treatment of neck pain. All patients received a tailored exercise program designed to challenge and improve neck strength except those with sinister pathology, radiological anomalies of the cervical or thoracic spine, or those suffering from migraine. Occasionally, and unpredictably, a patient was aggravated by the initial assessment procedures and experienced discomfort that did not settle within 48 hours. Exercise therapy was not encouraged for these patients because of the risk of aggravating symptoms. Outcomes to treatment were recorded at discharge.

Available data were divided into two samples. Sample 1 data were used to examine the relationship between presenting characteristics and response to therapy and to develop predictive equations. Sample 2 data were used to test the models developed on Sample 1.

Participants. Sample 1 was derived from 218 consecutive patients who were eligible for inclusion in the exercise program and who consented to the use of their de-identified data for research purposes. Of these, 122 started treatment. Reasons for not commencing treatment during these time frames were mostly personal ($n = 61$). In a few cases a patient did not meet the inclusion criteria or was aggravated by the assessment, moved, or commenced treatment in another state of Australia, preferred to wait on completion of legal proceedings, or was advised against the program by a doctor or third party. Sample 2 was derived from the next 309 consecutive patients presenting for treatment. Two hundred fourteen started treatment programs. Reasons for not commencing treatment were very similar to those reported for Sample 1.

Initial assessment. Prior to commencing the exercise program, all participants were assessed for history and duration of complaint, age, gender, and compensation status. Participants completed the Neck Disability Index (9). The Hanoun Multi-Cervical Unit (MCU, Hanoun Medical Inc., Toronto, Canada and Denver, Colorado) was used to measure range of motion (ROM) of cervical spine flexion, extension, left and right lateral flexion and left and right rotation.

To measure range of motion, participants were allowed familiarization trials until ROM was consistently produced and they reported being comfortable with the test procedures. Three consecutive test movements were then performed and averaged to provide a score for each movement.

The MCU does not permit assessment of rotation strength. Isometric strength of flexors, extensors, and lateral flexors was measured with the head in neutral and with the head rotated to 25 and then 45 degrees to the left and to the right. To measure strength, each participant was allowed 3-4 trials of graduated isometric effort and a final trial at maximum effort. Following this, 3 maximal isometric contractions were measured for flexion, extension, and lateral flexion. Each test was of 3-seconds duration and a 10-second rest occurred between tests. All patients received a standardized instruction to 'push as hard and as fast as you can' before each of the three test trials. The 3 maximal contraction trials were averaged to provide a single isometric strength measure (lbs) for each movement direction.

ROM values were compared to the values published by the AMA (10). Strength values were compared to isometric strength measurements of 100 healthy subjects (11).

Treatment. The exercise program was designed to challenge weak movements and achieve strength scores comparable to those reported for unimpaired subjects. Initial resistance to targeted movements was provided at 25-40% of the maximum isometric score achieved during testing. Participants attended 2-3 times per week for a 30 minute, supervised exercise program. At each session, they performed 3 sets of 10 repetitions for each of 6-8 exercises.

Resistance to exercise was progressed in response to participant perception of the effort required to repeat an exercise. At each session, participants rated perceived effort on a 9-point visual analogue scale during the first three repetitions of each movement. A score of 1 indicated that no effort was required, while a score of 9 indicated that the required effort was unacceptable. If the effort rating was 1-3, the resistance was progressed by 1lb. A rating of 4-6 incurred an increment of half a pound. The resistance was not altered for a rating of 7-8 and was lowered by half a pound for a rating of 9.

Participants were assessed for strength, range of movement, and NDI scores after every 9 sessions and were encouraged to continue the program until strength targets had been achieved.

Predictor variables. Variables examined for predictive utility were participant age, sex, duration of symptoms, compensation status, initial NDI scores, individual NDI item scores, and cervical spine range of movement and strength measurements.

Outcome. At discharge, participants were classified as responders if their NDI scores had changed by 14 points or more, and as non-responders if they changed by less than 14 points.

Data analysis. Bivariate logistic regression analysis was applied to determine variables that were significant predictors of response to exercise. Those variables that were significantly related to outcome were then entered into a multivariate logistic regression analysis. Logistic regression analysis assumes that predictor variables are measured on interval, ratio, or dichotomous scales (12). Individual items of the NDI are measured on an ordinal scale. Individual item scores were dichotomized at all possible cut points on the six-point scale, creating five new variables for each original item. Initial NDI scores were investigated as potential predictors both in continuous and dichotomized form. The process of dichotomizing NDI item and total scores resulted in multiple variables for each original variable. When more than one dichotomized variable belonging to a single raw score was a significant predictor, the variable with the highest overall classification accuracy was entered into multiple logistic regression.

The Wald statistic was used to determine significance of a predictor variable in a logistic regression analysis. Predictor variables were retained in a regression model only when the Wald statistic of each was significant with 95% confidence.

RESULTS

Sample 1 and 2 characteristics are summarized in Table 1.

Complete data sets were available for 97 subjects (79%) in Sample 1. Of these, 54 (55.6%) were classified as responders and 43 (44.3%) as

TABLE 1. Sample 1 and 2 characteristics

	Sample 1	Sample 2
n =	122 (female, 65%)	214 (female, 67%)
Age years (x, SD)	41.3 yrs (12.4)	40.5 yrs (12.4)
Private	71%	65%
Compensable	29%	35%
Mmonths of symptoms*	36 (13-98)	60 (22-132)
NDI scores (0-100)*	34 (24-46)	36 (26-52)
Program length (wks)*	6 (6-9)	8 (6-12)
Complete data sets	79%	90%

*median and inter-quartile range

non-responders. After intervention, responders improved (decreased) their NDI scores by a median of 22 (IQ 18-30), non-responders improved by a median of 6 (IQ 0-8).

Bivariate analysis. Age, sex, duration of symptoms or compensation status did not predict response to exercise. The initial NDI score was a significant predictor in continuous and dichotomous forms. The dichotomy at 30 points (0-29, 30-100) had the highest overall classification accuracy. Seven of the ten NDI items were significant predictors of response. These were lifting (dichotomized at 0-2, 3-5), reading (0-1, 2-5), concentration (0-1, 2-5), work (0, 1-5), driving (0-2, 3-5), sleeping (0-1, 2-5) and recreation (0-1, 2-5). For all NDI items, the higher the score (the higher the disability) the greater the probability of response. Of all strength and range of movement measurements, only extension strength in 45 degrees rotation was a significant predictor. For this variable, lower values corresponded with higher probabilities of response.

Table 2 provides a summary of significant models and their prediction accuracy. In the first column is the predictor variable. The second column shows the odds of responding to the program. The fifth column reports the accuracy in predicting response or non-response based on higher values for the predictor variable. For example, participants with lifting scores of 3 or more (indicating some difficulty associated with lifting) were 3.57 more likely to respond than participants with lifting scores of 0, 1 or 2. When this predictor was used to allocate a person to

TABLE 2. Bivariate logistic regression analysis summaries

Variable	OR*	95% CI*	Classification Accuracy (%)*
Lifting (0-1, 2-5)	3.57	1.53-8.32	66
Reading (0-1, 2-5)	3.58	1.49-8.61	66
Concentration (0-1, 2-5)	3.78	1.52-9.37	63
Work (0, 1-5)	2.72	1.05-7.06	62
Driving (0-2, 3-5)	3.35	1.31-8.59	60
Sleeping (0-1, 2-5)	2.65	1.16-6.05	62
Recreation (0-1, 2-5)	2.56	1.04-6.29	62
Initial NDI score	1.05	1.02-1.08	69
NDI dichotomized at 30 (0-29, 30-100)	5.56	2.23-13.85	70
Extension strength in 45° rotation	0.86	0.71-0.95	71

*OR = Odds ratio, 95% CI = 95% confidence interval of the odds ratio, Classification Accuracy (%) = the overall classification accuracy of the bivariate regression model

the category of responder or non-responder, the prediction was correct in 66% of cases.

Multivariate analysis. In multivariate analysis the model with the highest classification accuracy contained the two NDI items lifting (dichotomized at 0-1, 2-5) and reading (dichotomized at 0-1, 2-5). Odds ratios, confidence intervals, and overall classification accuracy for this model are shown in Table 3.

Comparisons of the overall classification accuracy between the multivariate model shown in Table 3 and the three best bivariate models (initial NDI score, initial NDI score dichotomised at 30, and extension strength in 45° rotation) had approximately equal accuracy. A summary of the resultant positive and negative predictive values for each model is presented in Table 4. This shows, for example, that an initial NDI score of 30% or more correctly classified 69.8% of responders and an NDI score of less than 30% correctly classified 70.6% of non-responders.

Testing the regression models on an independent sample. Complete data sets were available for 192 (90%) of Sample 2 patients who commenced the exercise program. Two outliers with exceptionally long treatment durations (beyond the 99th percentile) were removed from the analysis, leaving 190 subjects for testing the models. The only significant difference between the two samples was the duration of symptoms. Sample 2 participants had considerably longer history of neck pain, but outcomes of analysis were comparable when testing was repeated for the subgroup of Sample 2 with duration of symptoms matching those for Sample 1.

TABLE 3. Multivariate logistic regression summary

Variable	OR	95% CI	Class. Acc. (%)
NDI Lifting (0-1, 2-5)	2.77	1.14-6.73	67
NDI Reading (0-1, 2-5)	2.70	1.07-6.81	

TABLE 4. Positive and negative predictive values for four regression models

Model	+ve predictive values(%)	- ve predictive values(%)
NDI initial score	70	68
NDI dichotomized at 30	70	71
Extension strength in 45° rotation	73	67
Lifting (0-1, 2-5) and Reading (0-1, 2-5)	64	74

TABLE 5. Summary statistics of classification accuracy for each of the four regression models in both samples

Model	+ve predictive values%*		-ve predictive values%		Overall accuracy %	
	1	2	1	2	1	2
Sample						
NDI initial	70	55	68	62	69	57
NDI 30	70	55	71	65	70	58
Extension 45° rot.	73	54	67	65	71	55
Lifting (>1) and Reading (>1)	64	56	74	74	67	59

*+ve prediction: people who will respond to exercise

Sample 1 data indicated that approximately 56% of people appeared to respond to the program. Our models indicated that we might be able to predict between 64 and 73% of those who would respond. Similarly, Sample 1 data indicated that approximately 44% of people did not appear to respond to the program, and our models indicated that we might be able to predict between 67 and 74% of those who would not respond.

However, when we tested the models on a new sample, we found positive predictive values that ranged from 54-57%. These improved little or not at all on chance.

Our negative predictive models performed somewhat better, ranging from 62% (95% CI 49-72)-74% (95% CI 58-85) (Table 5).

DISCUSSION

This is the first study of which we are aware that has tested the utility of initial assessment characteristics in chronic neck pain patients for their utility in predicting response to an exercise program. Participants in this study had neck pain of 36 (Sample 1) and 57 (Sample 2) months average duration. Approximately half of each sample changed their NDI scores by more than 14 points (55.6% and 49.7% for Samples 1 and 2, respectively). Changes occurred in a relatively short period of time. The median treatment duration was 6 (IQ range 6-9) and 8 (6-12) weeks for Samples 1 and 2, respectively. Given the chronicity of the complaint, this is a promising proportion of patients. However, the program appeared to have little effect for approximately half of the participants. The characteristics that we are currently measuring do not provide us with adequate information to accurately predict the likeli-

hood of an individual with chronic neck pain responding to this program. This has important implications for the design of RCTs intended to determine the efficacy of exercise for individuals with chronic neck pain. Although substantial effects appear possible for large numbers of participants, effects will be diluted by a similar number of participants who may not derive benefit from exercise. We are unable to predict who will, and only weakly able to predict who will not, respond to exercise. Our RCTs may show us an average benefit of exercise, but will not advance our ability to identify those people who might benefit from exercise programs. The data observed in this study may account for the frustration anecdotally reported by clinicians when provided with guidelines based on the results of RCTs. At times, recommendations are made that a particular therapy is of little value, when clinicians see substantial responses in some individuals.

This study identified that items assessed using the NDI may have utility if incorporated into more sophisticated predictive models. For example, participants who indicated that they were able to lift and read with little discomfort appeared to benefit less from the program than people who were less comfortable with these activities. It is feasible that people who are already able to undertake challenging physical activities in their daily lives, without discomfort, may dominate the subgroup of people with a low likelihood of response to the physical challenge of an exercise program.

The data analyzed in this report were collected during the course of usual clinical care. We are now in an era where we have the technology to record and sort large volumes of measurements obtained during routine clinical practice. Useful analysis of this data becomes possible when a standardized approach to patient assessment is implemented. If clinicians agree on a minimal important data set that should be recorded for each patient entering a physical therapy program, data pooling and analysis could be part of usual care. There is considerable work to be undertaken in defining the items that might be productively included in a minimal important assessment and outcome data set for people presenting for treatment of chronic neck pain. Items identified as predictors of non-recovery following whiplash injury (13) might be considered in future studies. Hoving et al. (14) also identified a number of factors that are not measured by the NDI that may indicate response to exercise.

The predictive accuracy of our models diminished when tested on an independent sample. The majority of published studies of predictors of outcomes have not tested the utility of models on an independent sample. Our work indicates the important role of confirmatory studies

before the utility of predictors of outcomes can be endorsed by the scientific community.

KEY POINTS

- Approximately half of two samples of chronic neck pain participants in a strengthening program responded by a change in 14 NDI points.
- In the first sample, a prediction formula based on NDI scores and item responses predicted response or non-response with approximately 70% accuracy.
- When tested on the second sample, the same model predicted responders (positive predictive value 56%) and non-responders (negative predictive value 74%) with considerable loss of responder prediction accuracy.
- Being able to predict who is likely and unlikely to respond to a given therapeutic intervention has important practical implications for research design and clinical care.
- The prediction models derived in this study could be enhanced by the testing of additional predictor variables.
- The collection of data suitable for the testing and formulation of predictor models can occur in routine clinical care settings.

REFERENCES

1. Cote P, Cassidy D, Carroll L. The Saskatchewan health and back pain survey. *Spine* 1998;23:1689-98.
2. Verhagen AP, Peeters GCM, de Bie RA et al. Conservative treatment for whiplash (Cochrane Review). 2002 In: *The Cochrane Library*, Issue 1.
3. Hoving JL, Gross AR, Gasner D et al. A critical appraisal of review articles on the effectiveness of conservative treatment for neck pain. *Spine* 2001;26: 196-205.
4. Stratford PW, Riddle DL, Binkley JM et al. Using the Neck Disability Index to make decisions concerning individual patients. *Physiotherapy Canada* 1999;5: 107-112.
5. Bronfort G, Evans R, Nelson B et al. A randomized clinical trial of exercise and spinal manipulation for patients with chronic neck pain. *Spine* 2001;26:788-97.
6. Howell DC. *Statistical methods for psychology*. 3rd ed. Boston: PWS-Kent Pub. Co.; 1992.
7. Ylinen J, Takala E, Nykanen M et al. Active neck muscle training in the treatment of chronic neck pain in women. *JAMA* 2003;289:2509-16.

8. Korthals-de Bos IB, Hoving JL, van Tulder MW et al. Cost effectiveness of physiotherapy, manual therapy and general practitioner care for neck pain: Economic evaluation alongside a randomized controlled trial. *BMJ* 2003;326:911-6.

9. Vernon H, Mior S. The Neck Disability Index: A study of reliability and validity. *J Manipulative Physiol Ther.* 1991;14:409-15.

10. Cocchiarella L, Andersson G. Guides to the evaluation of permanent impairment. 5th ed. Chicago, Illinois, USA: AMA Press; 2000.

11. Jordan A, Mehlsen J, Bulow PM et al. Maximal isometric strength of the cervical musculature in 100 healthy volunteers. *Spine* 1999;24:1343-1348.

12. Menard, S. Applied logistic regression analysis. California: Sage Publications, Inc; 1995.

13. Scholten-Peeters G, Verhagen A, Bekkering G et al. Prognostic factors of whiplash-associated disorders: A systematic review of prospective cohort studies. *Pain* 2003;104:303-22.

14. Hoving JL, O’Leary EF, Niere KR et al. Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain* 2003; 102:273-281.



For FACULTY/PROFESSIONALS with journal subscription recommendation authority for their institutional library . . .

If you have read a reprint or photocopy of this article, would you like to make sure that your library also subscribes to this journal? If you have the authority to recommend subscriptions to your library, we will send you a free complete (print edition) sample copy for review with your librarian.

1. Fill out the form below and make sure that you type or write out clearly both the name of the journal and your own name and address. Or send your request via e-mail to getinfo@haworthpress.com including in the subject line "Sample Copy Request" and the title of this journal.
2. Make sure to include your name and complete postal mailing address as well as your institutional/agency library name in the text of your e-mail.

[Please note: we cannot mail specific journal samples, such as the issue in which a specific article appears. Sample issues are provided with the hope that you might review a possible subscription/e-subscription with your institution's librarian. There is no charge for an institution/campus-wide electronic subscription concurrent with the archival print edition subscription.]

YES! Please send me a complimentary sample of this journal:

(please write complete journal title here—do not leave blank)

I will show this journal to our institutional or agency library for a possible subscription.

Institution/Agency Library: _____

Name: _____

Institution: _____

Address: _____

City: _____ State: _____ Zip: _____

**Return to: Sample Copy Department, The Haworth Press, Inc.,
10 Alice Street, Binghamton, NY 13904-1580**