

Supplementary Information

Interactive Visualization of Patients Pain Trajectories and Chronicity Probability

For the visualization of all participants, see: <https://rstudio-connect.hu.nl/painr-app/>. In this visualization, "FALSE" indicates no chronic pain (pain < 3 at 6 weeks, 3 months, and 6 months), while "TRUE" denotes chronic pain (pain ≥ 3 at all time-points: 6 weeks, 3 months, and 6 months). The X-axis represents the pain score, measured using the Numerical Pain Rating Scale (0-10), and the Y-axis shows the cumulative number of days after the baseline measurement. "Patient_code" is a unique identifier for each patient. "LP" stands for linear predictor, "Prob" represents the probability of chronicity, and "Perc" indicates the percentual probability of chronicity. The bar graph and various values per variable illustrate the regression coefficient, multiplied by the patient data at baseline, across different variables from the prognostic model.



13 **Supplementary Information**14 **Appendix 1. TRIPOD Checklist Prediction Model Development and Validation**

15

Section/Topic	1	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5-6
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	5-6
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7-8
	5b	Describe eligibility criteria for participants.	7-8
	5c	Give details of treatments received, if relevant.	Not applicable
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	8
	6b	Report any actions to blind assessment of the outcome to be predicted.	7-8
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8-10
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	7-8
Sample size	8	Explain how the study size was arrived at.	10
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	10-11
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	10-11
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	10-11
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	10-11
Risk groups	11	Provide details on how risk groups were created, if done.	Not applicable
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	12-16
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	12-16
Model development	14a	Specify the number of participants and outcome events in each analysis.	13
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	17-18
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	17-20
	15b	Explain how to use the prediction model.	23-24
Model performance	16	Report performance measures (with CIs) for the prediction model.	19-22
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	28
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	25-28

Implications	20	Discuss the potential clinical use of the model and implications for future research.	28-29
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	30
Funding	22	Give the source of funding and the role of the funders for the present study.	30

16
17

18 **Supplementary Information**19 **Appendix 2 Table of contents**

20

21 **Link Github:**22 <https://github.com/uashogeschoolutrecht/painr>

23 Table of Contents Published with Bookdown:

24 **0. Introduction**

25 0.1 Data Flow

26

27 **1. Exploratory Data Analysis – Raw Data**

28 1.1 Suggested improvements of the code

29 1.2 Packages

30 1.3 Load data

31 1.4 First glimpse at missingness

32 1.5 Select relevant variables

33 1.6 Exploratory Data Analysis

34 1.7 Write table with all labels

35 1.8 Deal with ‘work’ variables

36 1.9 Recode physical_activity

37 1.10 Write subsetted data to disk

38

39 **2. Imputation of missing values**

40 2.1 Packages

41 2.2 Data

42 2.3 Prepare dataset for imputing

43 2.4 Convert all categorical vars to factors

44 2.5 Panel with all distributions

45 2.6 Imputation of missing values

46 2.7 Checking Missing Completely at Random (MCAR)

47 2.8 Missingness pattern

48 2.9 Define predictors to include in the imputations

49 2.10 Using the MICE package for imputation of missing values

50 2.11 Create predictorMatrix for MICE

51 2.12 Calculate percentage missing data and cases

52 2.13 Running the imputations

53 2.14 Inspect the imputations

54 2.15 Check convergence

55 2.16 Check for plausible values of imputation

56 2.17 Checking the used predictor matrix

57 2.18 Look at the datasets

58 2.19 Skimming the data

59 2.20 Add attitude

60 2.21 Save to disk

61

62 **3. Statistical exploration**

63 3.1 Packages

64 3.2 Data

65 3.3 Global parameters

66 3.4 Statistical analysis methods and missing data

67 3.5 Reformat dataframe to stacked format

68	3.6 Create time variable
69	3.7 Recode time
70	3.8 Adding baseline as time=0
71	3.9 Carry forward
72	3.10 Rework the graph above to get cumulative pain intensity scores
73	3.11 Get individual lines for each patient
74	3.12 Write to disk as excelfile and. Rds R binary file
75	3.13 Distribution of the data
76	3.14 Table: baseline characteristics of the included patients
77	3.15 Testing assumptions before backward analysis
78	
79	4. Prognostic model
80	4.1 Packages
81	4.2 Data load
82	4.3 Clean data and rename vars
83	4.4 Exploratory Data Analysis
84	4.5 Variable analysis – independent predictive capacity
85	4.6 Relevel dichotomous variables
86	4.7 Multivariable logistic regression analysis
87	4.8 AUC
88	4.9 Calibration curve
89	4.10 Result Calibration plot
90	4.11 Hoslem and Lemeshow
91	4.12 Model fit
92	4.13 Internal validation
93	4.14 Plot Corrected AUC
94	4.15 Results
95	4.16 Correcting the variables coefficients
96	4.17 Calibration in the Large
97	
98	5. Article figures
99	5.1 Packages
100	5.2 Data load
101	5.3 Clean data and rename variables
102	5.4 Exploratory Data Analysis
103	5.5 Variable analysis
104	5.6 Graph of model metrics
105	5.7 Relevel dichotomous variables
106	5.8 Refactor code above to a more compact version
107	5.9 Univariate analysis on the categorical variables
108	5.10 Visualize model outcome
109	5.11 Panel plot univariate
110	5.12 Multivariable logistic regression analysis
111	5.13 Backward model
112	5.14 Visualize backward model
113	5.15 Panel plot with all models
114	5.16 Rework figure labels
115	5.17 Figures for paper
116	5.18 Adding level info to figure
117	5.19 GGpubr panel
118	5.20 Area Under the receiver operating characteristic Curve (AUC)
119	5.21 Calibration curve

120 5.22 Formally testing the Goodness-of-fit using the Hosmer and Lemeshow

121 5.23 Intermezzo – linear predictors

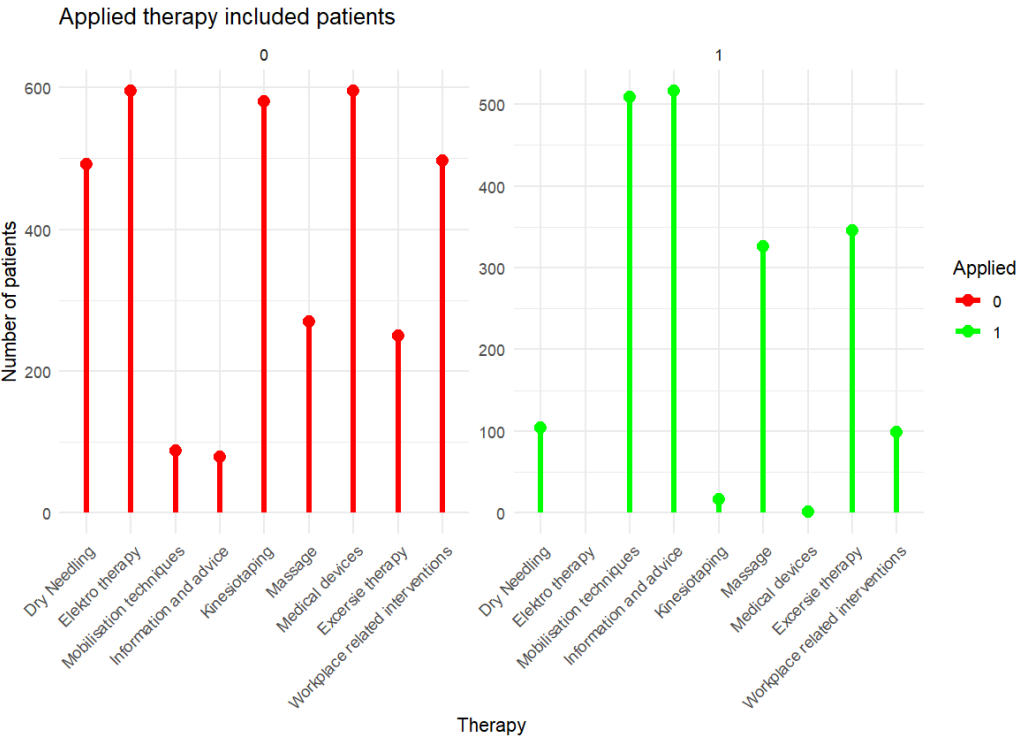
122

123 **Supplementary Information**

124 **Appendix 3 Overview Applied interventions study population**

125 **Table Intervention included patients (N = 596)**

Interventions	Number of patients	Applied (%)	Number of patients	Not applied (%)
1. Workplace, ergonomic and working time advice	99	16,6	497	83,4
2. Medical devices, collar or cervical pillow	1	0,2	595	98.2
3. Joint mobilizations, manipulation, traction, nerve mobilization techniques	509	85,4	86	14,6
4. Exercise therapy	346	58,1	250	41,9
5. Electrotherapy, laser, ultrasound, shockwave or heat therapy	0	0	596	100
6. Dry needling	492	17,4	104	82,6
7. Information and advice	79	86,7	517	13,3
8. Kinesiotaping	16	2,7	580	97,3
9. Massage	326	54,7	270	45,3



126

127 **Figure: Applied therapy included patients (N = 596)**

128