

---

# THE IMPACT OF PSYCHOLOGICAL TREATMENT ON CATASTROPHIZATION AND PHARMACOLOGICAL RESPONSE IN CHRONIC MIGRAINE: A SINGLE-CENTER EXPERIENCE

---

**Federica Nicoletta Sepe**  
Stroke Unit Department  
IRCCS C. Mondino Foundation  
Via Mondino, 2, 27100, Pavia, Italy  
[federica.sepe@mondino.it](mailto:federica.sepe@mondino.it)

**Claudia Lanni**  
Headache Clinic, Department of Neurology  
Azienda Sanitaria SS. Antonio e Biagio e Cesare Arrigo  
Via Venezia, 16, 15121, Alessandria, Italy  
[claudia.lanni@ospedale.al.it](mailto:claudia.lanni@ospedale.al.it)

**Daniele De Michelis**  
Headache Clinic, Department of Neurology  
Azienda Sanitaria SS. Antonio e Biagio e Cesare Arrigo  
Via Venezia, 16, 15121, Alessandria, Italy  
[daniele.michelis@ospedale.al.it](mailto:daniele.michelis@ospedale.al.it)

**Giacomo Lancia**  
Department of Basic and Applied Sciences for Engineering (SBAI)  
University of Rome "La Sapienza"  
Via Antonio Scarpa, 4. 00161, Roma.  
[giacomo.lancia@uniroma1.it](mailto:giacomo.lancia@uniroma1.it)

November 8, 2024

## ABSTRACT

**Background:** We aimed to assess the impact of pain catastrophizing, measured using the Italian version of the Pain Catastrophizing Scale (PCS), on the clinical response of patients with chronic migraine to anti-CGRP monoclonal antibodies combined with a multidisciplinary approach, including psychological treatment.

**Methods:** 25 Outpatients from SS. Antonio e Biagio e Cesare Arrigo headache clinic randomly assigned to receive Galcanezumab, Erenumab, or Fremanezumab. Their clinical response was evaluated over six months using various measures, including reducing the number of days with migraine per month, and quality of life using Headache Impact Test (HIT 6), MIgraine Disability Assessment Score questionnaire (MIDAS), and Beck's Inventory Scale (BDI II) scales to assess comorbid depression.

**Results:** We established a strong correlation between HIT 6 and PCS, with coefficients of 0.81 and 0.88 at  $T_1$  and  $T_2$ , respectively. Furthermore, we found no significant correlation between PCS and the other scales, such as MIDAS, as with any pharmacological therapies.

**Conclusion:** This study aims to clearly define the impact of a multidisciplinary approach including a psychological follow-up on a particular clinical phenotype of chronic migraines and their tendency to catastrophize, but more extended data are needed.

**Keywords** Psychological treatment; Chronic migraine; Catastrophization, Pain Catastrophizing Scale.

## 1 Introduction

The PCS is a 13-item self-report measure of catastrophizing in pain and a well-validated measure of maladaptive thinking patterns related to pain (Monticone et al., 2012). The PCS is composed of 3 subscales: rumination (e.g., “I can’t stop thinking about how much it hurts”), magnification (e.g., “I worry that something serious may happen”), and helplessness (e.g., “There is nothing I can do to reduce the intensity of the pain”) (Monticone et al., 2012). People affected by chronic headaches, defined according to the 3rd edition of the International Classification of Headache Disorders (ICHD III) (Arnold, 2018), represent a consistent group of patients suffering from chronic pain. Giving the best medical response to those patients, affording the considerable economic costs and psychological burden related to chronic pain, is a huge healthcare challenge. Since the 90s, researchers focused on the role of anxiety and mood disorders as comorbidities in migraine, without describing the specific role of a patient’s personality in the processing of pain (Pelzer et al., 2023a; Lipton et al., 2023; Klonowski et al., 2022). Catastrophizing can be considered a maladaptive cognitive response to a painful stimulation that influences negatively pain perception (Mathur et al., 2016). In high-frequency migraine patients, pain-related activity in the white matter structure of the insula correlates with pain catastrophizing and migraine severity (Kocakaya et al., 2023). Furthermore, in patients with medical overuse (MOH), a higher total PCS score was associated with decreased grey matter density in precentral and inferior temporal gyrus as an increased resting-state functional connectivity between middle temporal gyrus and cerebellum (Christidi et al., 2020). An abnormal reward mechanism dopamine-mediated can explain these structural changes: high-frequency migraine attacks induced sustained increases in dopaminergic trafficking that override homeostatic feedback control. This massive dopaminergic tone in reward could lead motivation and learning centres to actuate abnormal coping strategies such as catastrophizing for pain (Christidi et al., 2020; Maizels et al., 2012). Even if the role of pain-related cognitive processes and emotional state on pain-related disability is well established (Senturk et al., 2023a), how catastrophizing can influence therapeutical response in the calcitonin-gene related peptide (CGRP-mAbs) antibodies era is still debated. Interestingly, refractory migraine patients to CGRP- mAbs showed higher baseline PCS scores (Alvarez-Astorga et al., 2021) representing an independent negative predictor to CGRP-mAbs response (Silvestro et al., 2021; Mitsikostas et al., 2014).

This study aims to assess the potential impact of pain catastrophizing on the clinical response to CGRP-mAbs in a real-life setting at a tertiary headache centre in Northern Italy.

## 2 Materials and Methods

### 2.1 Ethics Committee

The local Clinical Research Ethics Committee approved the study (number of approval ASO.NEURO.21.02), and all patients signed informed consent forms

### 2.2 Study population

In this monocentric observational study, we enrolled 25 consecutive outpatients who visited the "SS. Antonio e Biagio e Cesare Arrigo" headache clinic between July 2021 and 2023. These were diagnosed with chronic migraine, with or without medication overuse, based on the ICDH III criteria. They were randomly assigned to receive one of three medications — Galcanezumab (120mg), Erenumab (140mg), or Fremanezumab (225mg) — following the local and EAN guidelines. The therapeutical response was assessed using two criteria: a  $> 50\%$  reduction in the frequency of number of days with migraine per month and a decrease in disability using the MIDAS (Stewart et al., 2001a) and the HIT 6 (Rendas-Baum et al., 2014) according to local and EAN guidelines. We also evaluated comorbid depression using the BDI II (BDI II  $> 13$ ) (Schotte et al., 1997a). Patients were assessed at the beginning of therapy ( $T_0$ ) and again at three months ( $T_1$ ) and six months ( $T_2$ ) after the start of treatment. Specialists in neurology and psychology treated each patient, and those with medical, psychiatric, or cognitive limitations, as well as those with chronic pain conditions or a history of substance abuse, were excluded from the study. The sociodemographic and clinical characteristics of patients are reported in Table 1.

Sex	Age	Ere nu mab	Fre ma ne zu mab	Gal ca ne zu mab	MID AS ( $T_0$ )	HIT 6 ( $T_0$ )	BDI II ( $T_0$ )	PCS ( $T_0$ )	Help ness ( $T_0$ )	Mag ni fi ca tion ( $T_0$ )	Ru mi na tion ( $T_0$ )	MID AS ( $T_1$ )	HIT 6 ( $T_1$ )	BDI II ( $T_1$ )	PCS ( $T_1$ )	Help ness ( $T_1$ )	Mag ni fi ca tion ( $T_1$ )	Ru mi na tion ( $T_1$ )	MID AS ( $T_2$ )	HIT 6 ( $T_2$ )	BDI II ( $T_2$ )	PCS ( $T_2$ )	Help ness ( $T_2$ )	Mag ni fi ca tion ( $T_2$ )	Ru mi na tion ( $T_2$ )
F	50-54		X		87	72	18	28	7	7	14	44.0	67.0	12.0	27.0	11.0	2.0	14.0	30.0	73.0	11.0	24.0	6.0	4.0	14.0
F	25-29		X		75	72	23	42	22	2	18	40.0	44.0	30.0	7.0	1.0	4.0	2.0	18.0	62.0	9.0	31.0	15.0	0.0	16.0
F	70-74		X		90	76	27	39	17	2	20	15.0	67.0	21.0	35.0	16.0	2.0	17.0	34.0	57.0	15.0	17.0	3.0	2.0	12.0
F	55-59		X		63	71	12	42	18	4	20														
F	20-24		X		160	65	30	33	15	2	16	43.0	60.0	7.0	29.0	11.0	0.0	18.0	58.0	59.0	4.0	22.0	9.0	0.0	13.0
F	45-49		X		76	44	10	26	8	0	18	80.0	54.0	8.0	24.0	8.0	0.0	16.0							
F	55-59	X			68	60	14	23	11	2	10	60.0	60.0	6.0	20.0	6.0	1.0	13.0	30.0	62.0	13.0	18.0	6.0	2.0	10.0
F	50-54	X			57	63	11	22	7	2	13	13.0	40.0	1.0	12.0	1.0	0.0	11.0	16.0	57.0	6.0	15.0	3.0	0.0	12.0
F	45-49	X			119	68	39	26	12	0	14	70.0	68.0	38.0	25.0	14.0	0.0	11.0	29.0	56.0	40.0	13.0	3.0	1.0	9.0
F	55-59	X			69	72	12	43	21	2	20														
F	40-44	X			97	67	17	22	8	0	14	29.0	50.0	15.0	16.0	4.0	2.0	10.0	9.0	36.0	6.0	3.0	0.0	1.0	2.0
M	65-69	X			30	63	23	32	12	0	20	30.0	46.0	11.0	11.0	1.0	1.0	9.0	12.0	36.0	11.0	8.0	0.0	1.0	7.0
F	65-69	X			12	60	3	24	14	2	8	11.0	46.0	9.0	15.0	6.0	2.0	7.0							
F	55-59			X	56	67	22	39	17	5	17	0.0	36.0	6.0	10.0	1.0	0.0	9.0	0.0	40.0	13.0	10.0	3.0	2.0	5.0
M	50-54			X	56	64	9	26	9	6	11	10.0	52.0	3.0	8.0	2.0	1.0	5.0	4.0	40.0	1.0	10.0	2.0	1.0	7.0
F	35-39			X	26	62	14	30	12	2	14	20.0	44.0	11.0	10.0	3.0	1.0	9.0	22.0	40.0	12.0	5.0	1.0	2.0	8.0
F	55-59			X	20	65	13	27	11	4	12	4.0	32.0	5.0	9.0	1.0	1.0	5.0	2.0	34.0	1.0	11.0	2.0	1.0	7.0
M	45-49			X	36	69	7	38	18	4	16														
F	60-64			X	87	62	11	35	15	4	16	12.0	62.0	10.0	26.0	10.0	3.0	13.0							
F	50-54			X	96	76	12	30	12	2	16														

Table 1: Demographic and clinical background details encompassing the entire patient cohort. Blank cells represent missing values. An X indicates the assignment of antibodies.

### 3 Statistical Analysis

As a primary outcome, we investigated the correlation between PCS and HIT 6 at  $T_0$ ,  $T_1$ , and  $T_2$  using Spearman's correlation test; see appendix A. We also examined the association between PCS and MIDAS and PCS and BDI II at these time points. A significance level of 5% was set for all hypotheses; with a Bonferroni correction, we set the p-value threshold at 0.00625.

We also assessed the reduction in MIDAS and HIT 6 using Wilcoxon's test, with the significance level adjusted to 0.00625; see appendix B.

Using the Jaccard Index, we investigated the similarity of subgroups of patients showing severe catastrophization (PCS > 30) as headache-related disability, considering HIT 6 > 50 and MIDAS > 11 separately. Depression at  $T_0$  was controlled using BDI II (> 13) as the antibody assigned; see C.

Using a logit approach, we explored the relationship between PCS at  $T_2$  and baseline information at  $T_0$ . The outcome variable was defined as a relative reduction of MIDAS > 50% between  $T_0$  and  $T_2$ . Permutation Importance identified the most impactful features for improving quality of life; see appendix D.

We focused on the primary outcome of PCS and HIT 6 correlation through Spearman's test to determine the sample size. Therefore, we utilized the method proposed by May and Looney (2020), setting a significance level of 0.0125 (adjusted for Bonferroni correction) and a power of 0.8; see appendix E. We estimated a sample size of 21, adjusted to 25 for a 15% dropout rate.

We used a simple randomization scheme to assign antibody therapy, ensuring equal assignment probabilities. Despite a higher number of patients receiving Galcanezumab, the chi-squared test confirmed no significant distribution difference among patients at a 5% significance level; see appendix F.

### 4 Results

The null hypothesis of no correlation between PCS and HIT 6 was rejected with p-values of 0.001 at  $T_0$ ,  $T_1$ , and  $T_2$ . Spearman's coefficients were 0.65, 0.81, and 0.88 at  $T_0$ ,  $T_1$ , and  $T_2$ , respectively. These results describe a strong and time-dependent correlation between PCS and HIT 6 scores. For PCS and MIDAS, the null hypothesis could not be rejected at  $T_0$  and  $T_1$  (p-value: 0.35 and 0.07) but was rejected at  $T_2$  (p-value < 0.001), with correlation coefficients of 0.81 and 0.68. So, as expected, a reduction in the tendency to catastrophize will directly impact on quality of life and number of attacks per month measured by MIDAS scale.

Significant differences were found for all scale scores between  $T_0$  and  $T_1$ : PCS (p-value < 0.001), HIT 6 (p-value < 0.001), and MIDAS (p-value < 0.001). Significant differences were also observed between  $T_0$  and  $T_2$ : PCS (p-value < 0.001), HIT 6 (p-value < 0.001), and MIDAS (p-value < 0.001). Figure 1 provides a detailed breakdown.

At the beginning of the study ( $T_0$ ), the Jaccard Index indicated a 47% agreement between the reduction of severe catastrophization scores and the other scales. This result indicates that 47% of patients with a high PCS score responded positively (the highest reduction in monthly migraine days) to both psychological and pharmacological treatments. This reduction occurred regardless of the type of antibody, except for a slight correlation with Galcanezumab treatment. Our model suggests that comorbid depression and antidepressant therapy at  $T_0$  have no influence.

The logit-based model had an AUROC of 0.75 (95% CI 0.73-0.78), i.e., one patient has a probability of 75% to have a good clinical response at six months from the beginning of the therapy. The logit-based model was inspected through the permutation importance method, revealing that age, Galcanezumab, and PCS score at  $T_0$  are the main factors in the clinical response, with importance scores of 0.29, 0.17, and 0.28 (95% CI 0.26-0.32; 0.15-0.18; 0.26-0.30), respectively.

Missing values resulted from right censoring, affecting six patients at either  $T_1$  or  $T_2$ . We handled the missing data by analyzing all available data; we excluded patients whose therapy was interrupted and for whom no further information was available.

### 5 Discussion

Our hypothesis posits that rumination, as a self-psychological condition, may directly influence the clinical efficacy of CGRP antibodies, particularly in cases deemed refractory. We propose that a multidisciplinary approach incorporating psychological support can significantly improve clinical response, especially for those exhibiting an inadequate response to treatment. However, identifying the specific patient phenotype that will benefit the most from this approach remains challenging. Surprisingly, patients with the best clinical response appear to be female and younger, with higher

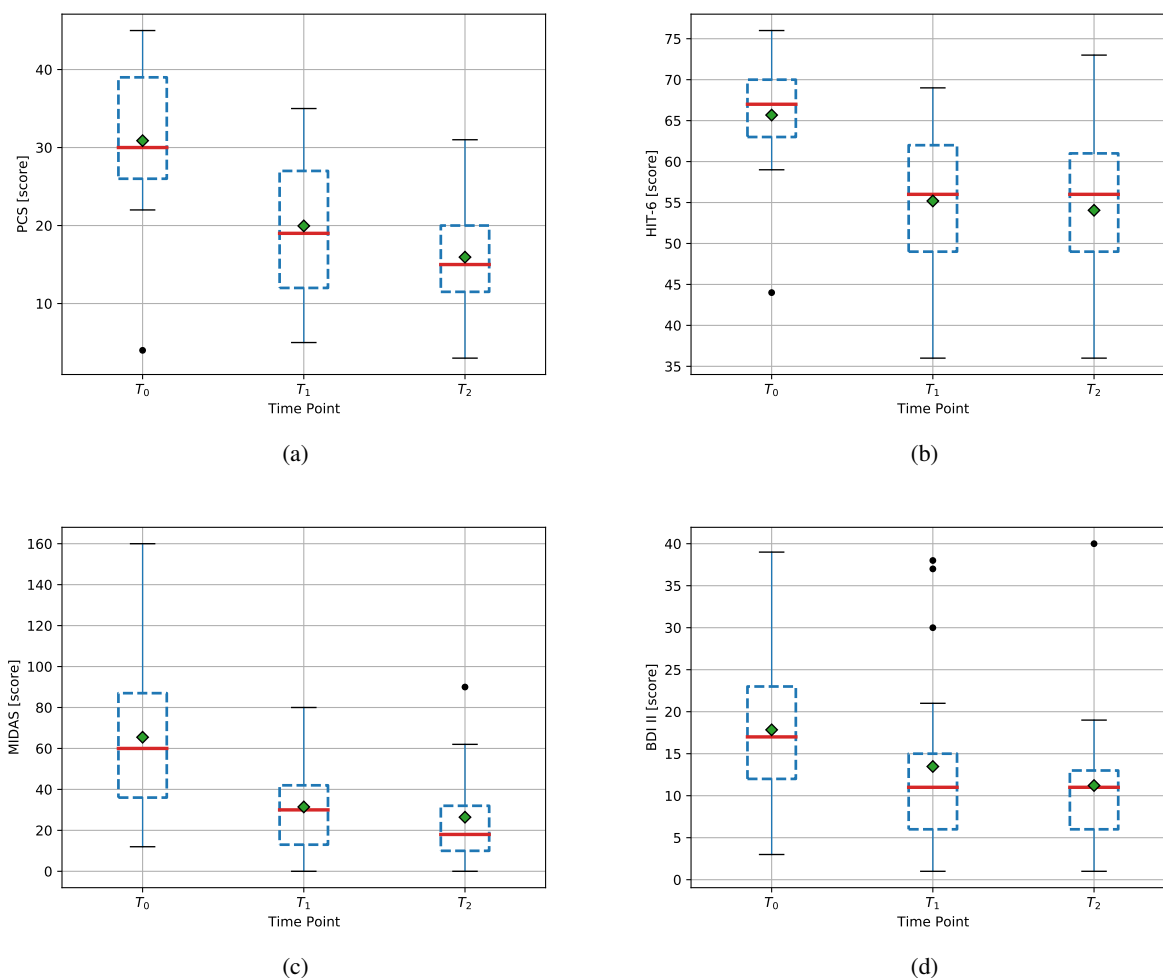


Figure 1: Panel with box plots for (a) PCS, (b) HIT6, (c) MIDAS, and (d) BDI II scales. The interquartile range is enclosed within the box, while the red dashed line is the median value. The green diamond is the average value. Black dots denote outliers.

PCS scores at baseline, regardless of comorbid depression. Galcanezumab seems to be more effective. There is a significant time-dependent correlation between the reduction in PCS and HIT 6 scores, particularly when longer medical and psychological therapies have been administered. This is independent of the number of attacks reported as of antidepressant previous therapies or depression. This suggests that a holistic treatment approach - combining pharmacological and psychological interventions - is more likely to improve the quality of life for chronic migraine patients. In a recent study, the HIT 6 score was found to have a weak correlation with PCS scores (Kim et al., 2021). However, no investigation was made into the correlation between any therapy and these variables. Another multicentric study also found that helplessness and anxiety are linked to the social quality of life of migraine patients when compared to controls. However, no correlations to ABS treatment were identified (Senturk et al., 2023b). We agree with previous theses that pain catastrophizing may indicate a clinical phenotype with heightened expression of altered central sensitization and consequential coping mechanisms leading to an overall decrease in quality of life (Sciricchio et al., 2019). Identifying the various factors that drive the progression of chronic migraine is crucial to developing effective prevention strategies (Driessen et al., 2022). Other real-world studies demonstrated that CGRP antibodies were able to exert an excellent clinical response across subgroups of migraine patients with comorbid psychological traits, mainly anxiety, and depression (Smitherman et al., 2020; Pelzer et al., 2023b). Unfortunately, this study has several limitations, mainly due to its observational monocentric design: the small sample size, the absence of a control group, and the short observation time. Further data of a more extensive nature is imperative to substantiate our findings.

	Mean	Variance	Skewness	Kurtosis	Tailing
PCS $T_0$	30	97	-0.58	3.75	-6.66
PCS $T_1$	20	49	0.11	1.95	0.38
PCS $T_2$	15	44	0.19	2.94	1.81
HIT6 $T_0$	65	21	0.45	2.55	2.96
HIT6 $T_1$	55	93	-0.43	2.26	-2.33
HIT6 $T_2$	54	110	-0.23	2.89	-0.63
MIDAS $T_0$	67	1135	0.79	3.43	6.29
MIDAS $T_1$	29	349	0.23	2.81	2.38
MIDAS $T_2$	26	557	1.32	3.42	9.11
BDI II $T_0$	20	92	0.38	2.62	2.34
BDI II $T_1$	14	118	1.05	3.16	5.81
BDI II $T_2$	11	75	1.38	7.48	23.91

Table 2: Descriptive statistics with Standardized moments for PCS, HIT 6, MIDAS, and BDI II at times  $T_0$ ,  $T_1$ , and  $T_2$ .

## 6 Conclusion

Treatment-related and individual factors contribute to the development of chronic migraine. A comprehensive approach is essential for individuals, encompassing both pharmacological and non-pharmacological interventions, as well as the management of behavioural and psychological factors. The development of personalized tools for predicting chronification represents a significant research priority. This study aims to delineate a clinical phenotype of chronic migraine patients characterized by a propensity to catastrophize. It advocates for a combined therapeutic strategy utilizing CGRP monoclonal antibodies alongside psychological counselling to enhance quality of life. Further investigation is necessary to assess the clinical implications of this integrative approach.

### A Spearman’s test

We utilized Spearman’s test (Spearman, 1961) to investigate the correlation between catastrophization (measured with PCS) and MIDAS, HIT 6, and BDI II scales. Specifically, we studied correlations at time  $T_0$  (starting of antibody treatment),  $T_1$  (3 months later), and  $T_2$  (six months later). For each couple of variables (i.e., PCS vs. MIDAS, PCS vs. HIT 6, and PCS vs. BDI II), we conducted a separate analysis.

Moreover, through the same approach, we studied the correlation level between the PCS subscales such as Helpness, Magnification, and Rumination, and the other scales mentioned above (i.e., MIDAS, HIT 6, and BDI II). Also, the correlation levels between the scales utilized and the antidepressant treatment at  $T_0$  were considered.

The significance level  $\alpha$  was set to 5%. However, the comparison of tests accomplished at the same time required the significance level to be adjusted in order to avoid the risk of a Type I Error. Thus, we utilized the Bonferroni correction (Dunn, 1961), i.e.,  $\alpha \rightarrow \frac{\alpha}{n}$ , with  $n$  the number of tests involved. Note that we are interested in testing correlation for two couples of variables along three distinct time points; so we have  $n = 8$ . As a result, the adjusted significance level is  $\bar{\alpha} = 0.00625$ .

For completeness, we report in Figures 2-4 the complete matrixes of the Spearman’s coefficients for each variable in our posess along with the p-values matrix; each matrix shows correlation levels at one of the three time-points  $T_0$ ,  $T_1$ , and  $T_2$ . The descriptive statistics of the scale scores are shown in Figure 2.

### B Wilcoxon’s Test

We employed Wilcoxon’s signed-rank test (Conover, 1999; McDonald, 2014) to assess significant differences in scale scores between  $T_0$  and subsequent time points, namely  $T_1$  and  $T_2$ . The scale considered is the same utilized in appendix A. Specifically, we conducted one-tailed tests for each pain scale to test the null hypothesis of no change in scale scores at  $T_1$  and  $T_2$  compared to  $T_0$ . The alternative hypothesis selected was a decrease in scale scores at  $T_1$  and  $T_2$  compared to  $T_0$ .

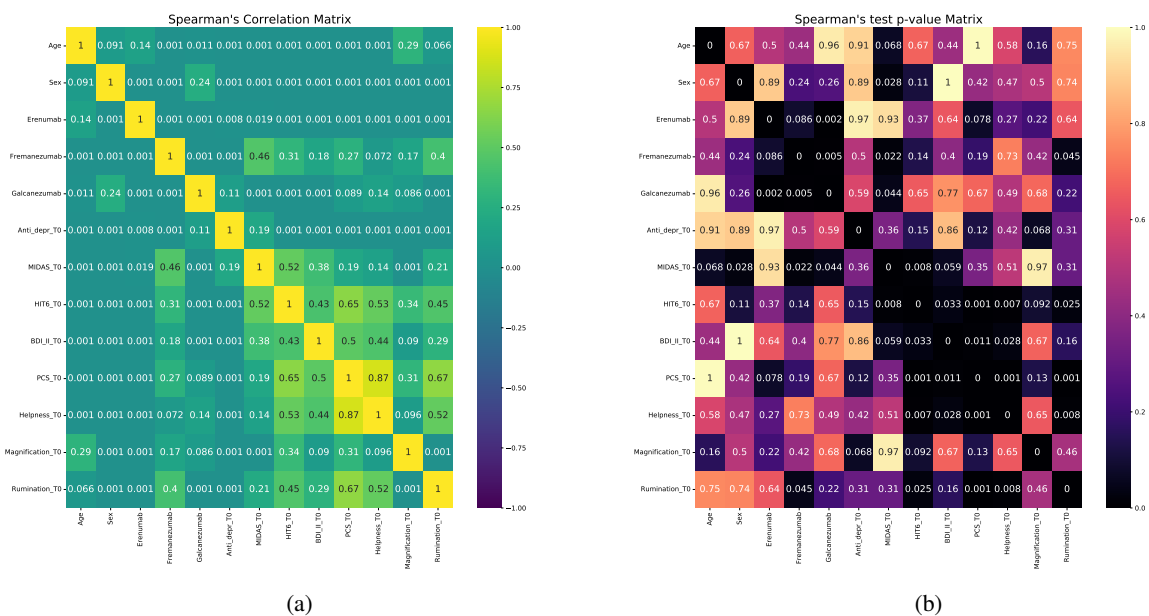


Figure 2: Spearman's correlation test at  $T_0$ . (a) Spearman's correlation coefficients and (b) p-values.

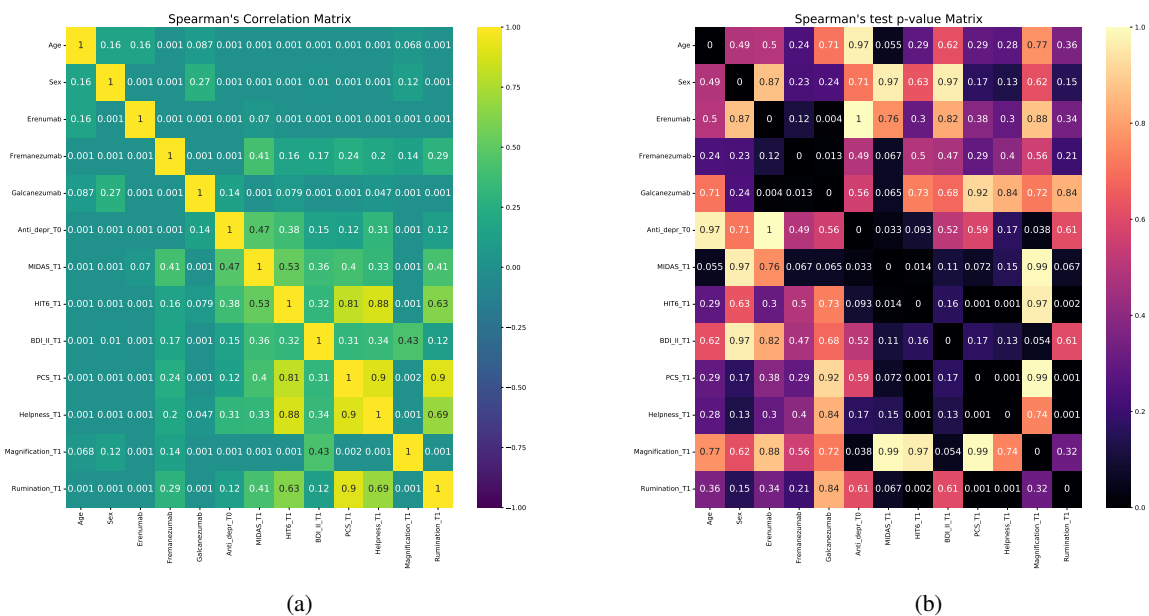


Figure 3: Spearman's correlation test at  $T_1$ . (a) Spearman's correlation coefficients and (b) p-values.



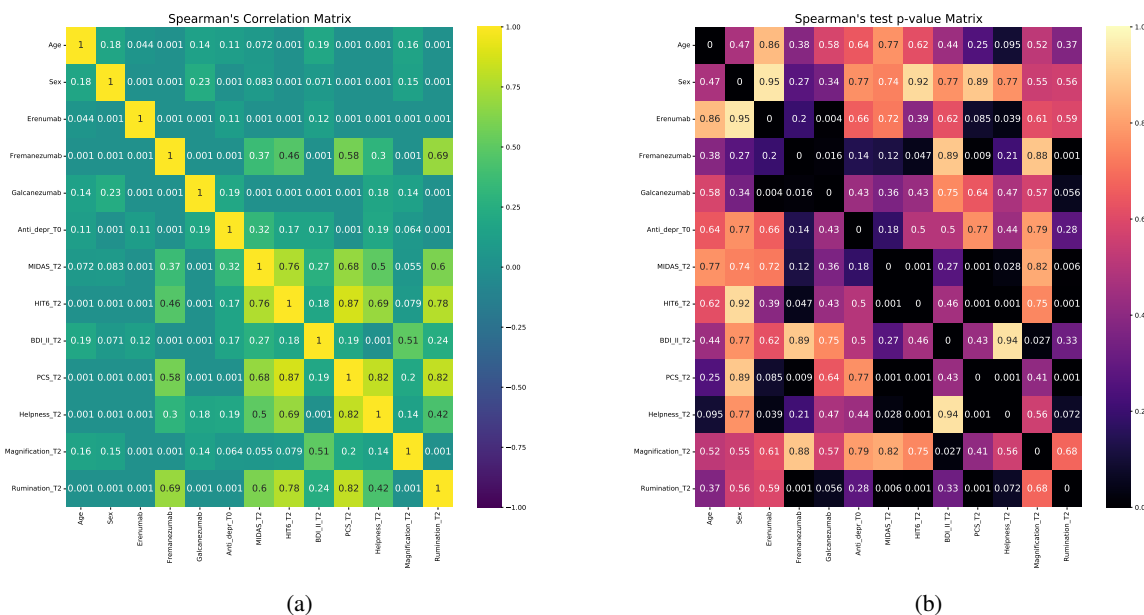


Figure 4: Sperman's correlation test at  $T_2$ . (a) Spearman's correlation coefficients and (b) p-values.

## C Jaccard Index

The Jaccard Index Chung et al. (2019) was employed to quantify the proportion of patients who exhibited both severe catastrophization and severe migraine-induced disability compared to those who did not.

To assess this relationship we defined a subgroup of patients characterized by severe catastrophization ( $PCS > 30$ ) and severe headache-related disability ( $HIT\ 6 > 50$ ,  $MIDAS > 11$ ). Thus, our primary goal was to determine whether patients prone to catastrophizing also experience severe disability due to migraines, and vice versa. We assessed the Jaccard Index across both the entire population and specific subpopulations derived through a stratification process. Stratification was performed based on either the antibody treatment received at  $T_0$  or by verifying the presence of comorbid depression ( $BDI\ II > 13$ , minimal depression).

To provide a clearer background about the Jaccard Index in this context, consider the following example. Let  $n$  denote the generic patient. Define  $y_{PCS}^{(n)}$  as a dichotomous variable that takes the value 1 if the patient's PCS score is greater than 30, and 0 otherwise. Similarly, let  $y_{MIDAS}^{(n)}$  be a dichotomous variable that takes the value 1 if the patient's MIDAS score is greater than 11, and  $y_{HIT6}^{(n)}$  take the value 1 if the patient's HIT 6 score is greater than 50. Also, let's consider the sets

$$\begin{aligned}\mathcal{Y}_{PCS} &= \{y_{PCS}^{(1)}, \dots, y_{PCS}^{(N)}\}; \\ \mathcal{Y}_{MIDAS} &= \{y_{MIDAS}^{(1)}, \dots, y_{MIDAS}^{(N)}\}; \\ \mathcal{Y}_{HIT-6} &= \{y_{HIT-6}^{(1)}, \dots, y_{HIT-6}^{(N)}\}.\end{aligned}$$

We denoted by superscript one patient within a generic set containing  $N$  individuals.

The Jaccard index for two sets  $A$  and  $B$  of binary items is denied as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}; \quad (1)$$

The Jaccard index measures similarity between binary sample sets, defined as the number of items with equal binary labels over the number of items considered.

Therefore,  $J(\mathcal{Y}_{PCS}, \mathcal{Y}_{MIDAS})$  informs about the percentage of patients which show either  $PCS > 30$  and  $MIDAS > 11$  or  $PCS < 30$  and  $MIDAS < 11$ . For both  $T_0$  and  $T_2$ , we evaluated  $J(\mathcal{Y}_{PCS}, \mathcal{Y}_{MIDAS})$ ,  $J(\mathcal{Y}_{PCS}, \mathcal{Y}_{HIT-6})$ , and  $J(\mathcal{Y}_{HIT-6}, \mathcal{Y}_{MIDAS})$ . Results are shown in Tables 3 and 4.



	No Stratification	Erenumab	Fremanezumab	Galcanezumab	Beck's score ( $T_0$ )
MIDAS - HIT6	0.94	0.99	0.99	0.99	0.93
MIDAS - PCS	0.47	0.21	0.75	0.51	0.57
HIT6-PCS	0.47	0.21	0.75	0.51	0.57

Table 3: Average Jaccard scores at time  $T_0$  for PCS, HIT6, and MIDAS. Each column reports average association indexes for each stratification (i.e., antibody or binarized Beck's score at  $T_0$ )

	No Stratification	Erenumab	Fremanezumab	Galcanezumab	Beck's score ( $T_0$ )
MIDAS - HIT6	0.52	0.61	0.99	0.03	0.56
MIDAS - PCS	0.05	0.01	0.25	0.01	0.05
HIT6-PCS	0.47	0.01	0.25	0.51	0.05

Table 4: Average Jaccard scores at time  $T_2$  for PCS, HIT6, and MIDAS. Each column reports average association indexes for each stratification (i.e., antibody or binarized Beck's score at  $T_0$ )

## D Logit-based approach

Using a logit approach involves modelling the relationship between patients' quality of life at  $T_2$  and their baseline information at  $T_0$ . The logistic regression model is typically used when the outcome variable is binary or categorical. In this case, the outcome variable  $Y$  is defined based on the relative reduction of MIDAS by more than 50% between  $T_0$  and  $T_2$ . This can be represented as:

$$Y = \begin{cases} 1 & \text{if } \frac{\text{MIDAS}_{T_0} - \text{MIDAS}_{T_2}}{\text{MIDAS}_{T_0}} > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Here,  $\text{MIDAS}_{T_0}$  and  $\text{MIDAS}_{T_2}$  denote the MIDAS scores at baseline  $T_0$  and follow-up  $T_2$ , respectively.

Permutation Importance (Breiman, 2001) is a technique used to identify the most impactful features in a predictive model. It works by evaluating the change in model performance (e.g., accuracy, AUC) when the values of a feature are randomly permuted while keeping other features unchanged. The decrease in model performance after permutation indicates the importance of that feature in predicting the outcome.

The logit model is a type of regression analysis used to predict the probability of a binary outcome based on one or more predictor variables. In practice, the probability that patients reduce significantly of 50% the value of MIDAS is given by

$$P(Y = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 X_1 - \dots - \beta_p X_p)}$$

where  $\beta_0$  is the intercept,  $\beta_1, \dots, \beta_p$  are the coefficients for predictors  $X_1 \dots X_p$ .

We used a 250-repeated 4-fold stratified cross-validator to evaluate the Area Under the Curve (AUC) and Brier's Score (BS) to assess the model's effectiveness. When utilizing the Permutation Importance method, we determined the resilience of each covariate by assigning importance scores based on the average difference of AUC just before and after permutating a covariate.

Note that with the term 250-repeated 4-fold stratified cross-validation approach we refer to the process we utilized to validate the model. It consists of repeating 250 times a 4-fold split into training and test datasets; usually, 3 folds are utilized to train the model, while the leftover is involved in the validation process. The model is therefore trained and tested through 250 different random configurations derived from the sample population. This method ensured robust assessment and minimized overfitting by repeatedly splitting the data into training and testing sets.

To determine the importance of each covariate, we used the Permutation Importance method. This involved calculating the importance scores by measuring the average change in AUC before and after permuting each covariate; see Figure 5 This approach allowed us to assess the resilience and significance of each covariate in predicting the outcome.

## E Sample Size

The estimation of the size was aimed to confirm the primary outcomes of our investigation. We recall that we were interested in exploring through Spearman's test a possible correlation between PCS and MIDAS as well as PCS and HIT 6 at time points  $T_0$ ,  $T_1$ , and  $T_2$ .

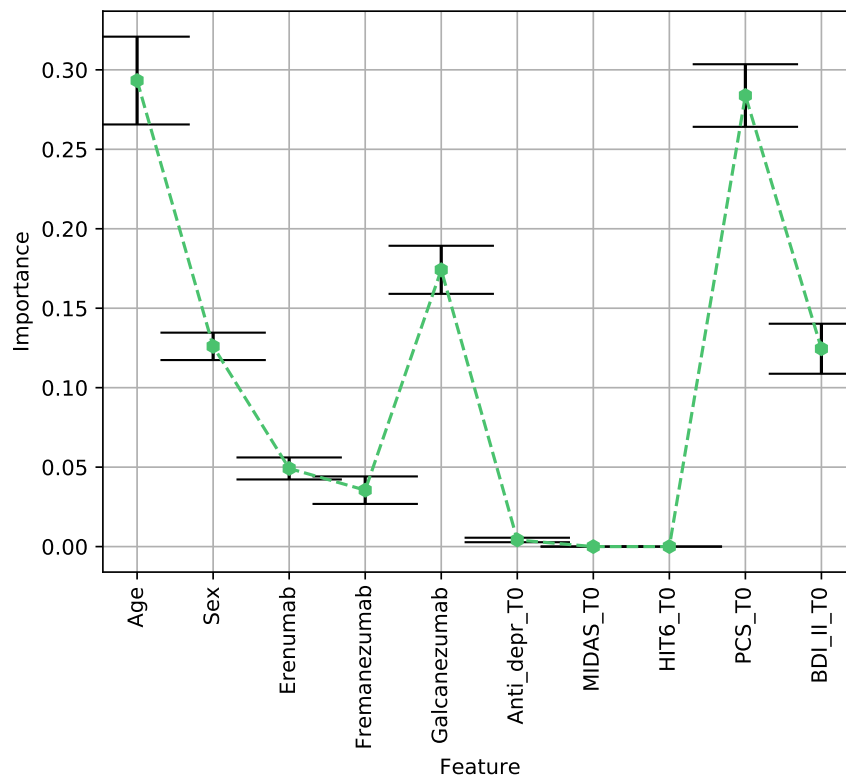


Figure 5: Mean Importance (with 95% CI) per each feature utilized in the logit prediction model. On the x-axis the features involved; note that the tick "MIDAS\_T0" refers to a binary variable derived by selecting MIDAS values greater than 11 at  $T_0$ . Likewise, "HIT6\_T0", "PCS\_T0", and "BDI\_II\_T0" refer to binary variables at  $T_0$  with cutoff, respectively, 50, 30, and 13. On the y-axis the mean importance values are reported; when appreciable, confidence bars are also reported.

The formulas provided by May and Looney (2020) represent the key to determining the required sample size for inference based on Spearman's test. To apply this method, it is required to specify the punctual values of both the null and alternative hypotheses. Therefore, we assumed to null hypothesis to be  $H_0 : \rho_0 = 0$ , and the alternative hypothesis  $H_0 : \rho_1 = 0.65$ . That is, we want to investigate no correlation ( $H_0 : \rho_0 = 0$ ) against a moderate (or even stronger) correlation, i.e.,  $\rho_1 = 0.65$ . Thus, the sample size  $n$  is given by

$$n = b + c^2 \left[ \frac{z_{\alpha/2} + z_{1-\phi}}{\Xi(\rho_0) - \Xi(\rho_1)} \right]^2, \quad (2)$$

where  $b$  and  $c$  are constants taking values 3 and 1, respectively;  $\alpha$  is the significance level and  $\beta$  the test's power; we used  $\Xi(\cdot)$  to denote the hyperbolic arctangent. Note that a regular choice of  $\alpha = 0.10$  needs to be adjusted to avoid incurring an increase of Type I errors due to multiple testing. To achieve this, we utilized the Bonferroni correction, so we adopted  $\alpha = 0.0125$ ; see appendix A. We also chose a high-power test to exclude the presence of Type II error, so we opted for  $\phi = 0.8$ .

We considered the possibility of dropouts during the phase of data acquisition. Thus, we adjusted the sample size through the following formula

$$\bar{n} = \frac{1}{1 - \delta}; \quad (3)$$

where  $\delta$  denotes the *dropout rate*; we opted for  $\delta = 0.15$ . Applying both (2) and (3) we obtained a population size of 25.

## F Randomization

We implemented a *simple randomization* Altman and Bland (1999) scheme to assign antibody therapy to participants. This method ensured that each participant had an equal probability of receiving the antibody therapy, thereby minimizing

selection bias and ensuring a balanced distribution of treatments across the study groups. As outlined in the main manuscript, Erenumab was assigned to 6 patients, Fremanezumab to 7 patients, and Galcanezumab to 12 patients. Despite assigning Galcanezumab to the majority of patients, the number of doses assigned resulted in being equally likely according to the chi-squared test.

We aimed to test the null hypothesis that the antibodies were assigned to patients with equal probability. As described, we utilized the chi-squared test with two degrees of freedom, setting a significance level of  $\alpha = 0.05$ . The test yielded a  $\chi^2$  statistic of 1.625 and a p-value of 0.4437. Given these results, we do not have sufficient evidence to reject the null hypothesis, indicating that the distribution of antibody assignments is consistent with equal probability.

## G Scales

MIDAS measures the headache-related disability within the last 3 months in three areas of life: work/school, household chores, and family/social/leisure activities. The MIDAS score is formed by summing the items and ranges from 0 to 270. The disability can be categorized into four grades: (I) little or no disability (MIDAS score 0–5), (II) mild (6–10), (III) moderate (11–20), and (IV) severe disability (> 21) (Stewart et al., 2001b).

HIT 6 is a six-item self-reported questionnaire used to assess headache-related disability. It assesses headaches' impact on psychological, cognitive, occupational, and social functioning over the previous four weeks. Scores range from 36 to 78, and scores above 60 indicate that headache seriously impacts functioning (Rendas-Baum and et al., 2014).

PCS was used to measure pain-related catastrophic thoughts. It consists of three subscales: helplessness (Qs 1-5 and 12), magnification (Qs 6, 7, and 13), and rumination (Qs 8-11). Evaluations (0 = not at all to 4 = always) are made using a five-point Likert scale. The PCS is rated from 0 to 52 points, and a higher score corresponds to a higher level of pain-related catastrophic thoughts (Darnall and et al., 2017).

The Beck Depression Inventory-II (BDI II) is a 21-item assessment of depression's severity. Each item is graded between 0 and 3 on a four-point Likert-like scale. Higher scores indicate more severe depressive symptoms (Schotte et al., 1997b).

## Acknowledgments

The authors would like to acknowledge the "Azienda Sanitaria SS. Antonio e Biagio e Cesare Arrigo" for their support.

## Conflicts of interest

The authors declare no conflicts of interest.

## Declaration of generative AI in scientific writing

The authors declare that no generative AI and AI-assisted technologies have been utilized during the writing process of this manuscript.

## References

- Altman, D. G. and Bland, J. M. (1999). Treatment allocation in controlled trials: why randomise? *Bmj*, 318(7192):1209–1209.
- Alvarez-Astorga, A. et al. (2021). Pain catastrophizing in a population of patients with migraine. *Neurologia*, 36(1):24–28.
- Arnold, M. (2018). Headache classification committee of the international headache society (ihs) the international classification of headache disorders. *Cephalalgia*, 38(1):1–211.
- Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.
- Christidi, F. et al. (2020). Dimensions of pain catastrophizing and specific structural and functional alterations in patients with chronic pain: Evidence in medication-overuse headache. *World Journal of Biological Psychiatry*, 21(9):726–738.
- Chung, N. C., Miasojedow, B., Startek, M., and Gambin, A. (2019). Jaccard/tanimoto similarity test and estimation methods for biological presence-absence data. *BMC bioinformatics*, 20(Suppl 15):644.

- Conover, W. J. (1999). *Practical nonparametric statistics*, volume 350. John Wiley & Sons.
- Darnall, B. D. and et al. (2017). Development and validation of a daily pain catastrophizing scale. *J. Pain*, 18:1139–1149.
- Driessen, M. T., Cohen, J. M., Thompson, S. F., Patterson-Lomba, O., Seminerio, M. J., Carr, K., Totev, T. I., Sun, R., Yim, E., Mu, F., et al. (2022). Real-world effectiveness after initiating fremanezumab treatment in us patients with episodic and chronic migraine or difficult-to-treat migraine. *Journal of Headache and Pain*, 23:56.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American statistical association*, 56(293):52–64.
- Kim, S., Bae, D. W., Park, S. G., and Park, J. W. (2021). The impact of pain-related emotions on migraine. *Scientific Reports*, 11(1).
- Klonowski, T., Kropp, P., Straube, A., and Ruscheweyh, R. (2022). Psychological factors associated with headache frequency, intensity, and headache-related disability in migraine patients. *Neurological sciences*, pages 1–12.
- Kocakaya, H., Say, B., Yörübulut, S., and Ergün, U. (2023). Emotion dysregulation in migraine patients: can it be a hallmark the probability of the transformation from episodic to chronic? *Neurological Research*, 45(7):610–618.
- Lipton, R. B., Buse, D. C., Nahas, S. J., Tietjen, G. E., Martin, V. T., Löff, E., Brevig, T., Cady, R., and Diener, H.-C. (2023). Risk factors for migraine disease progression: a narrative review for a patient-centered approach. *Journal of Neurology*, 270(12):5692–5710.
- Maizels, M., Aurora, S., and Heinricher, M. (2012). Beyond neurovascular: Migraine as a dysfunctional neurologic pain network. *Headache*, 52(10):1553–1565.
- Mathur, V. A., Moayed, M., Keaser, M. L., Khan, S. A., Hubbard, C. S., Goyal, M., and Seminowicz, D. A. (2016). High frequency migraine is associated with lower acute pain sensitivity and abnormal insula activity related to migraine pain intensity, attack frequency, and pain catastrophizing. *Frontiers in Human Neuroscience*, 10:489.
- May, J. O. and Looney, S. W. (2020). Sample size charts for spearman and kendall coefficients. *Journal of biometrics & biostatistics*, 11(2):1–7.
- McDonald, J. (2014). Wilcoxon signed-rank test. *Handbook of biological Statistics*, pages 186–189.
- Mitsikostas, D. D. et al. (2014). Refractory chronic cluster headache: a consensus statement on clinical definition from the european headache federation. *Journal of Headache and Pain*, 15:1–6.
- Monticone, M., Baiardi, P., Ferrari, S., Foti, C., Mugnai, R., Pillastrini, P., Rocca, B., and Vanti, C. (2012). Development of the italian version of the pain catastrophizing scale (pcs-i): cross-cultural adaptation, factor analysis, reliability, validity and sensitivity to change. *Quality of Life Research*, 21:1045–1050.
- Pelzer, N., de Boer, I., van den Maagdenberg, A. M., and Terwindt, G. M. (2023a). Neurological and psychiatric comorbidities of migraine: concepts and future perspectives. *Cephalalgia*, 43(6):03331024231180564.
- Pelzer, N., de Boer, I., van den Maagdenberg, A. M. J. M., and Terwindt, G. M. (2023b). Neurological and psychiatric comorbidities of migraine: Concepts and future perspectives. *Cephalalgia*.
- Rendas-Baum, R. et al. (2014). Validation of the headache impact test (hit-6) in patients with chronic migraine. *Health Quality Life Outcomes*, 12:4–13.
- Rendas-Baum, R. and et al. (2014). Validation of the headache impact test (hit-6) in patients with chronic migraine. *Health Qual. Life Outcomes*, 12:4–13.
- Schotte, C. K. W., Maes, M., Cluydts, R., De Doncker, D., and Cosyns, P. (1997a). Construct validity of the beck depression inventory in a depressive population. *Journal of Affective Disorders*, 46(2):115–125.
- Schotte, C. K. W., Maes, M., Cluydts, R., De Doncker, D., and Cosyns, P. (1997b). Construct validity of the beck depression inventory in a depressive population. *J. Affect. Disord.*, 46:115–125.
- Sciruicchio, V. et al. (2019). Pain catastrophizing in childhood migraine: An observational study in a tertiary headache center. *Frontiers in Neurology*, 10.
- Senturk, I. A., Askin Turan, S., Eyigurbuz, T., Senturk, E., and Kale Icen, N. (2023a). Pain-related cognitive processes, pain interference, and alexithymia in patients with primary headaches. *Cureus*, 15(5):e39688.
- Senturk, I. A., Askin Turan, S., Eyigurbuz, T., Senturk, E., and Kale Icen, N. (2023b). Pain-related cognitive processes, pain interference, and alexithymia in patients with primary headaches. *Cureus*, 15(5):e39688.
- Silvestro, M. et al. (2021). Refractory migraine profile in cgrp-monoclonal antibodies scenario. *Acta Neurologica Scandinavica*, 144(4):325–333.
- Smitherman, T. A., Tietjen, G. E., Schuh, K., Skljarevski, V., Lipsius, S., D’Souza, D. N., and Pearlman, E. M. (2020). Efficacy of galcanezumab for migraine prevention in patients with a medical history of anxiety and depression: A post hoc analysis of the phase 3, randomized, double-blind, placebo-controlled regain, and pooled evolve-1 and evolve-2 studies. *Headache*, 60:2202–2219.

Spearman, C. (1961). The proof and measurement of association between two things.

Stewart, W. F., Lipton, R. B., Dowson, A. J., and Sawyer, J. (2001a). Development and testing of the migraine disability assessment (midas) questionnaire to assess headache-related disability. *Neurology*, 56.

Stewart, W. F., Lipton, R. B., Dowson, A. J., and Sawyer, J. (2001b). Development and testing of the migraine disability assessment (midas) questionnaire to assess headache-related disability. *Neurology*, 56.