Group Report

Medical Application and Health-care
Nicolò Vella

Introduction

This project investigates the real-world effectiveness of anti-calcitonin gene-related peptide (CGRP) monoclonal antibodies (mAbs) in the treatment of chronic migraine. Utilizing a dataset from a prospective study of 179 subjects, the analysis focuses on the reduction of monthly migraine days (MMDs) across three consecutive one-year treatment cycles based on patient-reported diaries.
The cohort comprises individuals with chronic migraine (attacks >15 days/month), a condition associated with significant disability and comorbidities such as depression and anxiety. From a therapeutic perspective, patients were treated with one of three CGRP-targeting mAbs: erenumab (n=117), galcanezumab (n=53), and fremanezumab (n=9). Additionally, the dataset captures treatment history, noting that 31 patients had previously failed botulinum toxin treatment, while 87 had experienced a positive response.

Data Exploration

Baseline Data

Demographic and anthropometric measures such as AGE, WEIGHT, BMI, and INT_T0 show relatively smooth, unimodal distributions with limited outliers, suggesting good measurement quality. In contrast, several clinical-history variables (AGE_OF_ONSET, AGE_W_CHRONICMIGRAINE, MONTHS_OF_TREAT, and NUM_TRAT) are clearly right-skewed, indicating that most patients experienced early onset or shorter disease/treatment durations, with a smaller subset having long-standing or intensive treatment histories.

Longitudinal data

Overall, the categorical variables describe a heterogeneity in treatments, comorbidities, and symptoms, and a non-negligible amount of missing data (NA) across many features.
Treatment-related variables show that most patients have been exposed to at least one preventive or symptomatic therapy, reflecting the real-world prescribing patterns. Comorbidity-related features (Hypertension, Sleep_Disorders, Psychopathological) are mostly absent but present in a relevant minority, while FAMILIARITY is common, supporting a genetic predisposition. SEX is markedly imbalanced, with a predominance of females, consistent with migraine epidemiology

Demographic and Clinical Baseline Analysis

The demographic data confirms a female-predominant cohort (consistent with migraine epidemiology), with a slightly lower mean age for women (50.7 years) compared to men (53.1 years). Clinical history analysis reveals a strong genetic component, with 53% of patients reporting a family history of migraine. In terms of pharmacological history, patients show a "real-world" heterogeneity in previous treatments. Before starting monoclonal antibodies (mAbs), the largest proportion of patients (40%) utilized an association of different drugs, followed by combination drugs (30%). Regarding the current study treatment, the majority were assigned to Erenumab (117 patients), followed by Galcanezumab (53) and Fremanezumab (9).

Longitudinal Efficacy Analysis

The longitudinal data demonstrates a clear, positive therapeutic effect.

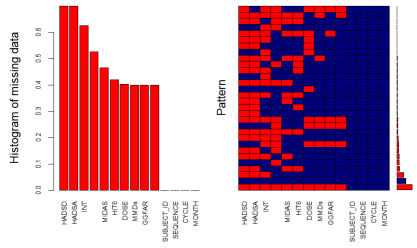
Frequency Reduction: There is a progressive decrease in Monthly Migraine Days (`MMDs`) across the three treatment cycles. The average `MMDs` dropped from a baseline of 14.38 days (Cycle 1) to 10.76 (Cycle 2) and finally 9.58 (Cycle 3).
This represents an average reduction of 4.8 days by the third cycle. The trend graph confirms that Cycle 3 consistently maintains the lowest trajectory of migraine days compared to previous cycles.
Intensity vs. Frequency: A crucial finding is the discrepancy between frequency and severity. While MMDs decreased linearly, the mean headache intensity did not show a corresponding downward trend. This suggests that while mAbs effectively reduce the number of attacks, the breakthrough attacks that do occur remain significant in pain severity.

Responder Rates and Discontinuation

Response rates indicate that the highest proportion of responders is observed in the first cycle with a Reduction of 61%.
Finally, the analysis of treatment discontinuation highlights that lack of therapeutic response was the overwhelming reason for dropouts. Notably, discontinuations due to Adverse Events (`AEs`) were the lowest category, suggesting the treatment is well-tolerated even among those for whom it is ineffective.

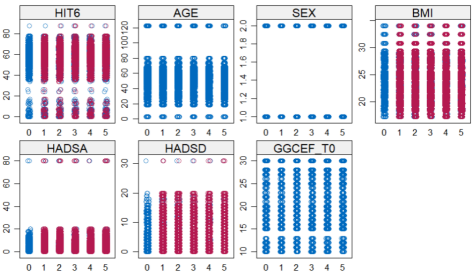
Missingness, Imputation and Analysis of the Data

Missing Data Analysis



Aggregation analysis reveals high missingness across clinical endpoints, with `HADSD/HADSA` exceeding 60% and `MIDAS/HIT6` between 40–50%. Conversely, structural identifiers remain fully observed. The data exhibits a non-random pattern of simultaneous missingness, suggesting unit non-response (attrition) rather than sporadic omission. Consequently, advanced imputation is required to avoid the bias inherent in complete-case analysis.

Imputation



After trying imputing the data with a baseline model that took the mean, (that we noticed to be not effective), we used the Library MICE (Multivariate Imputation by Chained Equations) that solves the missingness by treating each variable with missing data as a "target" to be predicted, using the other variables as predictors and as we can see in the image it worked following the paths of the data.

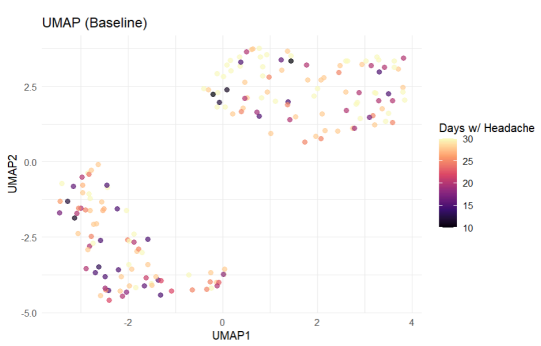
We decided to keep and impute all variables, and choose eventually to drop them if the task based on the data imputed would have poor quality results for the models.

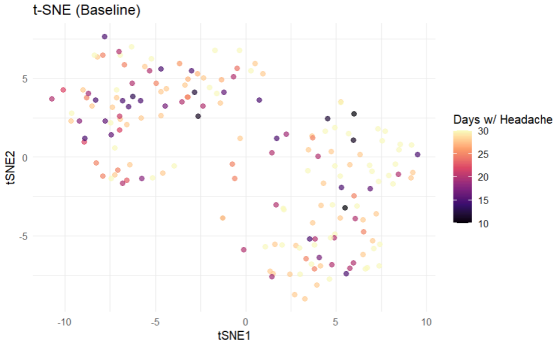
Analysis of the Imputed data

Analysis of baseline data

Both t-SNE and UMAP analyses reveal a distinct bimodal structure, separating the population into two major, disconnected clusters. Notably, disease severity does not drive this separation, as high-severity cases are distributed across both groups. Instead, this split suggests a dominant categorical variable

is the primary differentiator, likely reflecting biological sex or diagnosis type (e.g., Episodic vs.





Chronic).

Analysis of longitudinal data

The application of non-linear dimensionality reduction techniques, specifically t-SNE and UMAP, successfully reveals the latent topological structure of the clinical data, offering a superior representation to linear PCA by robustly handling outliers. The validity of this landscape is confirmed by a distinct visual gradient of HIT-6 scores, demonstrating that objective inputs like headache intensity is a strong predictor of subjective patient burden. Crucially, the UMAP projection uncovers a functional "trajectory of recovery," where successful treatment is observable as a physical migration of patient data points from severe, high-impact clusters at baseline toward mild, low-impact regions over time.