Medical Application and Healthcare - Module 1 Project Schema

Project Instructions

Students will work in ***groups of 2 to 4***. Each group must clearly describe the individual contributions and explain how the team collaborated to address the research questions.

The dataset description is provided below. Students should refer to Module 2 lessons to appropriately contextualize the clinical problem.

A set of ***example questions and related analytical tasks*** is provided as guidance. Students may select from the proposed questions, reformulate them, or propose original, clinically relevant questions.

A ***written report*** (maximum 4 pages) answering the selected questions must be submitted before the discussion session. Each project will be discussed in a ***10-minute presentation***. Groups may choose to support the discussion with slides.

Evaluation will focus on the coherence and consistency of the results, the quality of the discussion, and the student's understanding of how to appropriately apply complex analytical methods in clinical studies. Model performance will not be evaluated. However, students may be required to demonstrate the ability to interpret and justify poor model performance.

Analysis of Migraine Subjects' MMDs Changes and treatments response.

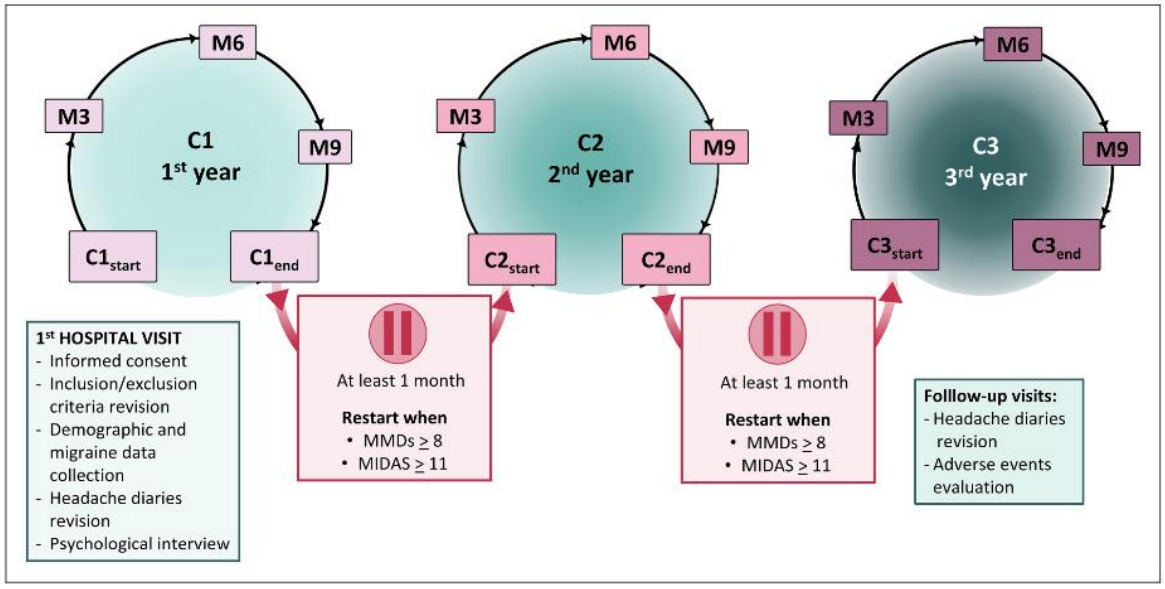
The dataset was collected as part of a prospective, real-world study aimed at assessing the effectiveness of anti-calcitonin gene-related peptide (CGRP) monoclonal antibodies (mAbs) across three consecutive one-year treatment cycles, with a primary focus on the reduction of monthly migraine days (MMDs).

A total of 179 subjects were enrolled (75.4% females; mean age 51.3 years, 95% CI [49.2–53.4]). Most participants (87.2%) had chronic migraine with or without medication overuse headache. All subjects initiated mAbs treatment between 2018 and 2020. Clinical data were collected prospectively using a structured headache diary throughout the three treatment cycles.

The original study results are published in: Vaghi G et al. Real-world effectiveness of anti-CGRP monoclonal antibodies over three consecutive one-year treatment cycles:

An intention-to-treat analysis. Cephalalgia. 2025 Aug;45(8):3331024251353421. doi: 10.1177/03331024251353421. Epub 2025 Aug 6. PMID: 40770917.

Here a schema of the study design. Legend: C1 = first one-year cycle of treatment; C2 = second one-year cycle of treatment; C3 = third one-year cycle of treatment; MMDs = monthly migraine days; MIDAS = Migraine Disability Assessment scale; M0 = baseline; M3 = three months from treatment initiation or re-initiation; M6 = six months from treatment initiation or re-initiation; M9 = nine months from treatment initiation or re-initiation.



Below, the two dataset files are described. Missing data are indicated with NA values.

Data set description – Baseline

SUBJECT ID	Unique Patient Identifier				
AGE	Age in years				
SEX	1 female	2 male			
DIAGNOSIS	1 chronic migraine	2 medication overuse headache	3 high-frequency episodic migraine		
Suspension	0 3-year treatment period completed	1 treatment discontinued before 3 years			
MONTHS_OF_TREAT	Number of months of treatment				
TREATMENT_DISC	Reason for treatment discontinuation				
AGE_OF_ONSET	Age of migraine onset				
AGE_W_CHRONICMIGRAINE	Age with chronic migraine				
FAMILIARITY	1 at least one first-relative with migraine	0 no first-relative with migraine			
WEIGTH	Weight in Kg				
HEIGTH	Height in m				
BMI	Body mass index				
SIDE side if headache	1 right	2 left	3 bilteral	4 side shifting	
PULSATING pulsating quality of pain	1 yes	2 no			
PAIN_MOVMENT pain aggravated by movement	1 yes	2 no			
Aura presence of aura	1 yes	2 no			
T0_SYMPT_TREATMENT Type of acute drug used as baseline	1 non-steroideal anti-inflammatory drugs	2 triptans	3 association of different drugs	4 combination drugs	
ANTIBODY Type of anti-CGRP monoclonal antibodies	1 erenumab	2 galcanezumab	3 fremanezumab		
Bbloc Failure of B-blockers	1 yes	2 no			
Caant Failure of calcium antagonists	1 yes	2 no			
Tricyclic	1 yes	2 no			

Failure of tricyclic antidepressants					
Antiepil Failure of anti-epileptic drugs	1 yes	2 no			
SSRISNRI Failure of antidepressant like SSRI and SNRI	1 yes	2 no			
Antiipnt Failure of anti-hypertensive drugs	1 yes	2 no			
Botulin Failure of botulinum toxin	1 yes	2 no			
DETOXPRE Previous detoxification protocol for medication overuse headache	1 yes	2 no			
Psycopathological Concomitant psycopathological disorders	1 yes	2 no			
Hypertension Concomitant arterial hypertension	1 yes	2 no			
Sleep_Disorders Concmiptant sleep disorders	1 yes	2 no			
PREV_T0 Ongoing oral prevention at baseline	1 yes	2 no			
GGCEF_T0 Number of monthly headache days at baseline					
INT_T0 Severity of headache at baseline	0 to 10				

Data set description – Longitudinal

SUBJECT ID Unique Patient Identifier					
CYCLE Treatment Cycle	1-3	Note: each cycle was separated by a 3-month mAbs suspension period			
MONTH Month within each Cycle	1, 3, 6, 9, 12				
MMDs Days with Migraine per month					
RED_MMD_VST01 Changes in MMD from Visit 1					
DOSE Total number of doses of acute drugs per month					
GGFAR Days with intake of at least one analgesic drug per month					
HADSA HOSPITAL ANXIETY QUESTIONNAIRE	Score of HADS anxiety section				
HADSD HOSPITAL DEPRESSION QUESTIONNAIRE	Score of HADS depression section				
HIT-6 Headache Impact Test	Score of HIT-6 scale				
INT Severity of headache	1-10				
MIDAS MIDAS Disability scale	Grade 1 0-5	Grade 2 6-10	Grade 3 11-20	Grade 4 >21	

Examples of Research Questions and Tasks

Q1 – Tasks 0 and 1 – Descriptive Analyses and Preprocessing

- What is the average reduction in monthly migraine days (MMDs) from baseline across three consecutive one-year treatment cycles?
- Are treatment effects consistent across cycles?
- Do baseline MMDs progressively decrease over time?
- What proportion of patients achieve ≥30% and ≥50% reduction in MMDs at the end of each cycle and across the full treatment period?

Task 0 – Data Exploration

- Inspect baseline distributions (age, sex, diagnosis, comorbidities).
- Explore longitudinal trends in MMDs and headache severity.
- Quantify dropouts and treatment discontinuations.

Tasks 1 - Missing Data Imputation (Baseline & Longitudinal Data)

- Explore missingness patterns and data characteristics (e.g., monotonicity) in baseline variables (e.g., demographics, disease history) and longitudinal variables (e.g., repeated measurements by cycle and month).
- Compare and select appropriate imputation strategies for baseline variables, including mean/median imputation, regression imputation, and most frequent category imputation for categorical variables.
- Compare and select appropriate model-based imputation strategies for baseline variables, particularly multivariate imputation methods (e.g., MICE). Based on the research questions, compare imputation methods while also considering the requirements and assumptions of the specific analytical model.
- Assess imputation strategies for longitudinal variables by comparing different methods, including Last Observation Carried Forward (LOCF), Next Observation Carried Backward (NOCB), and linear interpolation.

Q2 – Tasks 2 – Predictors of Response and Discontinuation

- Which baseline patient characteristics (e.g., demographics, comorbidities, migraine history) predict treatment response?
- Which factors are associated with treatment discontinuation?

Tasks 2 — Mixed-Effect Model Analysis

- Fit baseline models to assess the association between baseline characteristics and treatment response.
- Fit linear mixed-effects models with Outcome: Monthly Migraine Days (MMDs), Fixed effects: time, cycle, month, and baseline severity, Random effects: patient-specific intercepts and, where appropriate, patient-specific slopes
- Evaluate MMD reduction from baseline to the final month of follow-up and Fit models with different temporal granularities and estimate changes in MMDs over time (by cycles and months) using a hierarchical modeling framework.

Q3 - Tasks 3 —Migraine Phenotypes and Disease Trajectories

- Do distinct patient phenotypes (e.g., stable responders, late responders, fluctuating burden, non-responders) emerge from longitudinal data?
- Do MMDs temporal patterns emerge from data and are they predictive of treatment response?
- Can these (computational) phenotypes be used to stratify patients for better treatment outcomes?

Tasks 3 — Temporal Data Mining and Electronic Phenotyping

- Identify distinct disease progression patterns across patients using unsupervised approaches.
- Align time-series data and apply time-series binning to simplify individual trajectories. For each bin, compute representative statistics such as the mean, trend slope, and volatility.
- Convert continuous monthly measurements into fewer, interpretable temporal intervals and apply temporal abstraction techniques to derive higher-level features.
- Interpret the resulting phenotypes and communicate them using meaningful visualizations.

Q4 - Tasks 4 — Survival Analysis

- What is the time to achieving a ≥50% reduction in MMDs?
- How do baseline covariates or phenotypes influence time-to-response and treatment discontinuation?

Tasks 4 — Application of Kaplan Maier, Log Rank Test and Survival regression models

- Define the response event of interest (e.g. first month with ≥50% reduction in MMDs from baseline OR conclusion of treatments cycles) and construct survival dataset with time-to-event variable, event indicator, baseline covariates
- Fit Kaplan–Meier curves (overall and stratified by phenotype AND/OR cycle), apply the correct test and check the assumptions
- Fit Cox proportional hazards models to estimate hazard ratios for predictors of response/discontinuation. Check proportional hazards assumptions and visualize hazard ratio forest plots.