Prediction using machine learning of cumulative drug use for neuropathic pain treatment and its determinants in people with diabetic neuropathy in a national cohort

First published: 22/07/2026

Last updated: 22/07/2026

Study Planned

Administrative details

EU PAS number

EUPAS1000001018

Study ID

1000001018

DARWIN EU® study

No

Study countries

Spain

Study description

Observational, longitudinal, retrospective cohort study based on real-world data from BIFAP. The study population will include adults aged 18 years or older with an incident recorded diagnosis of diabetic neuropathy between 1 January 2019 and 31 December 2021. Each patient will have an individualised three-year follow-up from the first recorded diagnosis. The main objective is to develop and validate a predictive model to estimate, from baseline patient characteristics, the three-year cumulative defined daily dose of drugs used for neuropathic pain treatment, specifically tricyclic antidepressants, serotonin-norepinephrine reuptake inhibitors and α2δ subunit ligands. Secondary objectives are to identify baseline predictors by drug family and to characterise longitudinal patterns of registered pharmacological use. Candidate predictors will include demographic, clinical, laboratory, pharmacological and healthcare utilisation variables available at or before the index date. Analyses will be conducted in R and Python using reproducible workflows, including gradient boosting approaches with Tweedie loss and, as an alternative approach, multitask hurdle neural network models. Model performance, sensitivity analyses and interpretability using SHAP or DeepSHAP will be assessed.

Study status

Planned

Research institutions and networks

Institutions

CSEU La Salle

Spain

First published: 03/06/2026

Last updated: 03/06/2026

Institution | Educational Institution

Contact details

Study institution contact

Miguel Molina-Álvarez miguel.molina@lasallecampus.es

Study contact

miguel.molina@lasallecampus.es

Primary lead investigator

Miguel Molina-Álvarez 0000-0002-3368-9962

Primary lead investigator

ORCID number:

0000-0002-3368-9962

Study timelines

Date when funding contract was signed

Planned: 13/05/2026

Actual: 13/05/2026

Study start date

Planned: 15/06/2026

Data analysis start date
Planned: 01/10/2026

Date of final study report
Planned: 28/02/2027

Sources of funding

More details on funding

Academic study promoted by CSEU La Salle, Universidad Autónoma de Madrid.

Study protocol

Protocolo.pdf (216.55 KB)

Regulatory

Was the study required by a regulatory body?

No

Is the study required by a Risk Management Plan (RMP)?

Not applicable

Other study registration identification numbers and links

http://doi.org/10.17605/OSF.IO/5TP8Q

Methodological aspects

Study type

Study type list

Study topic:

Other

Study topic, other:

Prediction modelling of cumulative pharmacological use in diabetic neuropathy using real-world data.

Study type:

Non-interventional study

Scope of the study:

Drug utilisation

Method development or testing

Data collection methods:

Secondary use of data

Study design:

Retrospective longitudinal cohort study using BIFAP data. Adults with incident diabetic neuropathy recorded between 2019 and 2021 will be followed for three years from first diagnosis to model cumulative registered pharmacological use.

Main study objective:

To develop and validate a predictive model to estimate, from baseline patient characteristics, the three-year cumulative defined daily dose of drugs used for neuropathic pain treatment in patients with diabetic neuropathy, specifically tricyclic antidepressants, serotonin-norepinephrine reuptake inhibitors and α2δ subunit ligands.

Study Design

Non-interventional study design

Cohort

Study drug and medical condition

Medicinal product name, other

The pharmacological exposure of interest will include the following active substances and ATC codes: N02BF01 gabapentin, N02BF02 pregabalin, N06AX21 duloxetine, N06AX16 venlafaxine, N06AX23 desvenlafaxine, N06AA02 imipramine, N06AA04 clomipramine, N06AA06 trimipramine, N06AA09 amitriptyline, N06AA10 nortriptyline and N06AA12 doxepin. Some active substances were not available as selectable INN taxonomy terms in the catalogue interface and are therefore specified using their corresponding ATC codes.

Study drug International non-proprietary name (INN) or common name

GABAPENTIN

PREGABALIN

DULOXETINE

VENLAFAXINE

DESVENLAFAXINE

CLOMIPRAMINE HYDROCHLORIDE

AMITRIPTYLINE

Anatomical Therapeutic Chemical (ATC) code

(N02BF01) gabapentin

gabapentin

(N02BF02) pregabalin

pregabalin

(N06AX21) duloxetine

duloxetine

(N06AX16) venlafaxine

venlafaxine

(N06AX23) desvenlafaxine

desvenlafaxine

(N06AA02) imipramine

imipramine

(N06AA04) clomipramine

clomipramine

(N06AA06) trimipramine

trimipramine

(N06AA09) amitriptyline

amitriptyline

(N06AA10) nortriptyline

nortriptyline

(D04AX01) doxepin

doxepin

Medical condition to be studied

Diabetic neuropathy

Population studied

Short description of the study population

Adults aged 18 years or older included in BIFAP with an incident recorded diagnosis of diabetic neuropathy between 1 January 2019 and 31 December 2021. The study will be conducted in Spain using routinely collected real-world data from BIFAP, mainly from longitudinal primary care electronic health

records and, when available, complementary hospital information. The index date will be defined as the date of the first recorded diagnosis of diabetic neuropathy during the inclusion period, and each patient will be followed for three years from that date. Patients younger than 18 years, patients with active cancer, women pregnant during follow-up, patients with predefined active psychiatric disorders, patients with insufficient information to define essential study variables, and patients with incomplete follow-up preventing adequate longitudinal observation will be excluded. No sampling will be performed; all eligible patients available in BIFAP who meet the selection criteria will be included.

Age groups

- **Adult and elderly population (≥18 years)**
  - Adults (18 to < 65 years)
    - Adults (18 to < 46 years)
    - Adults (46 to < 65 years)
  - Elderly (≥ 65 years)
    - Adults (65 to < 75 years)
    - Adults (75 to < 85 years)
    - Adults (85 years and over)

Study design details

Setting

The study will be conducted in Spain using routinely collected real-world data from BIFAP. The cohort will include adults aged 18 years or older with an incident recorded diagnosis of diabetic neuropathy between 1 January 2019 and 31 December 2021. The index date will be the first recorded diagnosis of

diabetic neuropathy within the inclusion period. Each patient will have an individualised three-year follow-up from the index date. Patients younger than 18 years, patients with active cancer, women pregnant during follow-up, patients with predefined active psychiatric disorders, patients with insufficient information to define essential variables, and patients with incomplete follow-up preventing adequate longitudinal observation will be excluded. No treatment arms or comparators are defined, as this is a non-interventional retrospective cohort study.

Comparators

No formal comparator group is defined. This is a non-interventional retrospective cohort study. The study will compare levels and longitudinal patterns of registered pharmacological use across patients according to baseline demographic, clinical, laboratory, pharmacological and healthcare utilisation characteristics.

Outcomes

The primary outcome will be three-year cumulative pharmacological use, expressed as cumulative defined daily dose, of drugs used for neuropathic pain treatment: tricyclic antidepressants, serotonin-norepinephrine reuptake inhibitors and α2δ subunit ligands. This outcome will be calculated from BIFAP prescription records, overall and by drug family, and will be interpreted as registered pharmacological use rather than actual consumption or adherence. Complementary longitudinal indicators may include treatment initiation, persistence, discontinuation, treatment escalation and switching between pharmacological families, depending on data availability and quality.

Data analysis plan

Analyses will be conducted in R and Python using predefined reproducible workflows. Descriptive analyses will summarise baseline characteristics,

outcome distribution, missing data and correlations among predictors. Candidate predictors will be restricted to variables available at or before the index date to avoid look-ahead bias. Data will be split into training, validation and independent test sets, with preprocessing, imputation and encoding performed within closed pipelines to avoid data leakage. The reference model will be XGBoost or gradient boosting with Tweedie loss, suitable for skewed continuous outcomes with possible excess zeros. An alternative multitask hurdle neural network may be developed to model separately the probability and magnitude of pharmacological exposure. Performance will be assessed using MAE, RMSE and R^2 for continuous outcomes, and AUC-ROC, precision-recall metrics, sensitivity, specificity and calibration for binary hurdle components. Interpretability will be assessed using SHAP and, where appropriate, DeepSHAP. Sensitivity analyses will evaluate the robustness of analytical decisions.

Data management

ENCePP Seal

The use of the ENCePP Seal has been discontinued since February 2025. The ENCePP Seal fields are retained in the display mode for transparency but are no longer maintained.

Data sources

Data source(s)

BIFAP - Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito Público (Pharmacoepidemiological Research Database for Public Health Systems)

Data sources (types)

Drug prescriptions

Electronic healthcare records (EHR)

Use of a Common Data Model (CDM)

CDM mapping

No

Data quality specifications

Check conformance

Unknown

Check completeness

Unknown

Check stability

Unknown

Check logical consistency

Unknown

Data characterisation

Data characterisation conducted

No