

ESTRUCTURA RECOMENDADA Y CONTENIDO DEL PROTOCOLO¹

1. Título del estudio

Predicción mediante aprendizaje automático del consumo acumulado de fármacos para el tratamiento del dolor neuropático y sus determinantes en personas con neuropatía diabética en una cohorte nacional

2. Promotor del estudio: *nombre, dirección y datos de contacto.*

Promotor académico del estudio: CSEU La Salle, Universidad Autónoma de Madrid.

Investigador Principal: Miguel Molina Álvarez.

Dirección institucional y datos de contacto: Calle de la Salle 10, 28023 (Madrid) / miguel.molina@lasallecampus.es

3. Responsables del estudio: *nombres, títulos, grados, especialidad, lugar de trabajo y direcciones de todos los responsables, incluyendo el investigador coordinador, con una lista de todas las entidades colaboradoras primarias y demás ubicaciones pertinentes del estudio.*

IP: Miguel Molina Álvarez (MMA), Grado en Fisioterapia, Máster Universitario en Aspectos Básicos y Clínicos del Dolor, Máster en Ciencia de Datos y Doctor en Ciencias de la Salud por la URJC. Personal docente e investigador en CSEU La Salle, Universidad Autónoma de Madrid.

Jorge Menéndez Cámara (JMC), Grado en Fisioterapia, Máster en Estadística Aplicada y doctorando en Ciencias de la Salud por la URJC. Personal predoctoral en URJC.

Santiago Joaquín Tena Benítez (STB), Grado en Fisioterapia y alumno de Máster en Fisioterapia avanzada en el Tratamiento del Dolor en CSEU La Salle, Universidad Autónoma de Madrid.

¹ Según ANEXO I - Real Decreto 957/2020, de 3 de noviembre.

Luis Matesanz García (LMG), Grado en Fisioterapia, Máster en Terapia Manual Ortopédica en el Tratamiento del Dolor, Máster en Farmacología para Fisioterapeutas y Doctor en Ciencias de la Salud por la URJC. Personal docente e investigador en CSEU La Salle, Universidad Autónoma de Madrid.

4. **Resumen:** *resumen del protocolo del estudio, que tendrá las siguientes partes:*

a) Título y Autoría **Título del estudio:** Predicción mediante aprendizaje automático del consumo acumulado de fármacos para el tratamiento del dolor neuropático y sus determinantes en personas con neuropatía diabética en una cohorte nacional. **Versión y fecha:** V1 de 25 de marzo de 2026. **Autor principal:** Miguel Molina Álvarez **Organismo:** CSEU La Salle, Universidad Autónoma de Madrid.

b) Justificación y Contexto La neuropatía diabética (ND) es la neuropatía periférica más común. Afecta hasta al 50% de los pacientes con diabetes, presentándose dolor neuropático en el 20-40% de los casos. Esta condición impacta severamente la calidad de vida y se asocia a un aumento del uso de recursos sanitarios. Dado que los tratamientos actuales muestran una eficacia limitada y requieren seguimientos crónicos, existe una necesidad crítica de identificar predictores basales que permitan anticipar trayectorias de mayor consumo farmacológico registrado. El uso de datos masivos y aprendizaje automático permite captar la complejidad clínica y asistencial de esta progresión en condiciones de práctica clínica real más allá de los ensayos clínicos convencionales.

c) Hipótesis y Objetivos **Hipótesis:** Hipótesis: Las características basales individuales, incluyendo variables demográficas, clínicas, polimedicación y de utilización sanitaria, se asocian significativamente con la dosis diaria definida (DDD) acumulada a tres años de los fármacos utilizados en el tratamiento del dolor neuropático y permiten identificar patrones diferenciados de utilización farmacológica registrada. **Objetivo principal:** Desarrollar y validar un modelo predictivo para estimar, a partir de las características basales de los sujetos, la DDD acumulada a tres años de los fármacos utilizados en el tratamiento del dolor neuropático, concretamente de antidepresivos tricíclicos, inhibidores de la recaptación de serotonina y noradrenalina y ligandos de la subunidad $\alpha 2\delta$. **Objetivos secundarios:** Identificar factores predictores específicos por familia de fármacos y caracterizar los patrones de utilización farmacológica longitudinal registrada durante el periodo de seguimiento.

d) Diseño del Estudio: Se realizará un estudio de cohortes longitudinal retrospectivo basado en la base de datos nacional BIFAP. La cohorte incluirá pacientes con diagnóstico incidente registrado entre 2019 y 2021, con un seguimiento individualizado de tres años para cada paciente a partir de la fecha inicio.

e) Población Criterios de inclusión: Se incluirán adultos de 18 o más años con diagnóstico de ND entre 2019 y 2021. Se excluirán los menores de edad, mujeres embarazadas durante el seguimiento, los pacientes con enfermedad oncológica activa durante los años de estudio, aquellos con diagnósticos psiquiátricos graves previos que puedan condicionar de forma sustancial los patrones de prescripción, y los casos con falta de datos esenciales para el seguimiento o para la definición de variables clave del estudio.

f) Variables: La variable de resultado será la DDD acumulada a tres años de fármacos utilizados en el tratamiento del dolor neuropático. Este consumo farmacológico se operativizará como la suma acumulada de DDD registradas de cada familia terapéutica, durante los tres años posteriores a la fecha de inicio. Además, se explorarán de forma complementaria indicadores longitudinales de utilización, incluyendo inicio de tratamiento, persistencia, discontinuación, escalada terapéutica y cambio entre familias farmacológicas. Las variables predictoras incluirán datos sociodemográficos, clínicos, farmacológicos y asistenciales disponibles en el momento basal, tales como edad, sexo, IMC, HbA1c, colesterol, exposición farmacológica previa, polimedicación y variables de utilización de servicios sanitarios registradas en BIFAP.

g) Fuentes de los Datos: La fuente de información será la base de datos BIFAP, integrando registros longitudinales de Atención Primaria para el seguimiento de los pacientes y, cuando estén disponibles, datos hospitalarios para la identificación de diagnósticos y covariables clínicas en 12 Comunidades Autónomas.

h) Tamaño del Estudio: Se incluirán todos los sujetos disponibles en la base de datos que cumplan los criterios de selección. Dado el carácter poblacional y predictivo del estudio, el tamaño muestral vendrá determinado por el volumen de datos elegibles disponible, priorizando la precisión, estabilidad y validación de los modelos desarrollados, más que un cálculo clásico de potencia orientado a contraste de hipótesis.

i) Análisis de los Datos: El análisis se realizará en R y Python, comparándose distintos modelos predictivos adecuados a la distribución del resultado. Se incluirá un modelo de referencia basado en Gradient Boosting con función de pérdida Tweedie, especialmente útil ante distribuciones asimétricas y potencial exceso de ceros en el consumo farmacológico acumulado; así como un modelo alternativo de red neuronal multi-tarea con estructura hurdle para estimar simultáneamente la probabilidad y la magnitud del consumo. El rendimiento de los modelos se evaluará mediante procedimientos de validación interna y métricas apropiadas para variables continuas no normalmente distribuidas. La

interpretabilidad se abordará mediante valores SHAP y DeepSHAP para identificar la contribución relativa de los principales predictores.

j) Etapas y Calendario: Abril – Julio 2026: Aprobación por Comité de Ética y acceso a datos. Agosto – Septiembre 2026: Extracción, limpieza y estructuración del dataset. Octubre – Noviembre 2026: Análisis descriptivo e informes de situación. Diciembre 2026 – Enero 2027: Modelización y análisis de sensibilidad. Febrero 2027: Elaboración del informe final y manuscrito científico.

5. Modificaciones y actualizaciones: *toda modificación sustancial del protocolo del estudio después del inicio de la recogida de los datos, con su justificación, la fecha y la indicación de la sección del protocolo que se ha visto afectada.*

Toda modificación sustancial del protocolo del estudio que se produzca tras su aprobación inicial será documentada de forma explícita, incluyendo su justificación, la fecha de la modificación y la sección del protocolo afectada. En caso de no producirse modificaciones relevantes durante el desarrollo del estudio, se dejará constancia expresa de ello en la versión final.

6. Etapas: *cuadro con el calendario previsto para, al menos, las siguientes etapas:*

Etapa / Hito	Actividades Principales	Cronograma
a) Inicio de la recogida de datos	Aprobación BIFAP, envío a Comité de Ética, autorizaciones de acceso y definición final de variables.	1 Abril – 31 Julio 2026
b) Final de la recogida de datos	Extracción de base de datos, limpieza, validación y estructuración del dataset analítico.	1 Agosto – 30 Septiembre 2026
c) Informes de situación	Análisis descriptivo inicial, caracterización de la cohorte y evaluación de datos faltantes.	1 Octubre – 30 Noviembre 2026
d) Informes intermedios	Modelización predictiva (ML/Regresión), análisis de tendencias y análisis de sensibilidad.	1 Diciembre 2026 – 31 Enero 2027

e) Informe final de resultados	Interpretación de resultados, redacción del informe final y envío del manuscrito científico.	1 Febrero – 28 Febrero 2027
---	---	--

7. Justificación y contexto:

La neuropatía diabética (ND) es una complicación microvascular y neurológica de la diabetes tipo 1 y tipo 2 que afecta a los nervios periféricos, y puede presentarse también en personas con prediabetes. Esta complicación supone una importante carga clínica, asociándose a alteraciones de la marcha, pérdida del equilibrio y dolor neuropático. En fases avanzadas puede contribuir al desarrollo de úlceras, dificultad en la cicatrización, pie diabético y, en casos extremos, derivar en amputaciones. Dentro de este espectro, la ND dolorosa representa un problema particularmente incapacitante, con una prevalencia estimada entre el 20% y el 40% de las personas con diabetes. El dolor neuropático impacta de manera profunda en la calidad de vida: limita la movilidad, interfiere con el sueño, se asocia a síntomas depresivos y aumenta el uso de recursos sanitarios. Las personas con neuropatía dolorosa presentan mayor sufrimiento psicológico y peor bienestar que aquellos con neuropatía indolora o diabetes sin complicaciones, y en los casos de mayor intensidad, el dolor llega a dificultar o interrumpir la actividad laboral.

Desde el punto de vista terapéutico, la evidencia disponible muestra que los tratamientos farmacológicos actuales, incluyendo en primera línea los antidepresivos tricíclicos, los inhibidores de la recaptación de serotonina y noradrenalina y los ligandos de la subunidad $\alpha 2\delta$ (gabapentinoides), presentan una eficacia limitada. Los números necesarios para tratar suelen situarse por encima de 4–7 pacientes, y los perfiles de seguridad dificultan su mantenimiento a largo plazo. A pesar de las revisiones sistemáticas más recientes, las recomendaciones terapéuticas apenas han variado en la última década, lo que refleja la persistente dificultad para controlar de forma adecuada el dolor neuropático diabético. En este contexto, el consumo de medicación analgésica entre las personas con ND suele ser crónico, creciente y heterogéneo, respondiendo tanto a la evolución clínica individual como a factores demográficos, comórbidos, farmacológicos, sociales y organizativos del sistema sanitario. Sin embargo, pese a la elevada carga de enfermedad, existe escasa evidencia longitudinal poblacional que permita identificar qué características basales predicen un mayor consumo futuro o cómo evolucionan las trayectorias de uso farmacológico.

En la ND, conocer qué factores individuales se asocian al consumo farmacológico aporta un valor clínico y sanitario relevante. La identificación de predictores basales permitiría detectar, desde fases tempranas, a pacientes con mayor riesgo de incremento del consumo farmacológico en los años siguientes y, con ello, priorizar intervenciones dirigidas a optimizar las estrategias terapéuticas, intensificar medidas no farmacológicas y reducir efectos adversos evitables. Asimismo, el análisis de las tendencias de consumo permitirá distinguir patrones de utilización persistentes, crecientes o transitorios, lo que puede aportar información relevante sobre la evolución clínica, la respuesta terapéutica y el impacto de intervenciones o cambios en la práctica asistencial durante el seguimiento. En conjunto, la identificación de predictores individuales y la caracterización temporal del consumo podrían facilitar un abordaje más personalizado, mejorar la asignación de recursos y generar evidencia útil para el desarrollo de guías clínicas y políticas de seguridad farmacológica.

8. Hipótesis y objetivos de la investigación: objetivos primarios y secundarios.

Hipótesis: Las características basales individuales de los pacientes con ND, incluyendo variables demográficas (edad y sexo), clínicas y analíticas (comorbilidades, índice de masa corporal y parámetros analíticos de rutina), polimedicación y de utilización sanitaria previa al inicio de la cohorte (frecuencia de visitas), se asociarán con el consumo farmacológico acumulado, estimado mediante la dosis diaria definida (DDD) acumulada a tres años, de los fármacos utilizados en el tratamiento del dolor neuropático, tanto de forma global como por familias farmacológicas, incluyendo antidepresivos tricíclicos, inhibidores de la recaptación de serotonina y noradrenalina y ligandos de la subunidad $\alpha 2\delta$. Asimismo, estas variables basales permitirán identificar patrones diferenciados de trayectoria longitudinal de utilización farmacológica registrada durante los tres años posteriores al diagnóstico. De forma exploratoria, los modelos de aprendizaje automático permitirán identificar, jerarquizar y comparar predictores adicionales de la exposición acumulada y de las trayectorias de utilización farmacológica no especificados a priori.

Objetivos: Objetivo principal: Desarrollar y validar un modelo predictivo para estimar, a partir de las características basales de los sujetos, la DDD acumulada a tres años de los fármacos utilizados en el tratamiento del dolor neuropático, concretamente de antidepresivos tricíclicos, inhibidores de la recaptación de serotonina y noradrenalina y ligandos de la subunidad $\alpha 2\delta$. Objetivos secundarios: 1) Identificar los factores basales asociados a la exposición farmacológica acumulada por familias farmacológicas de interés durante los tres años posteriores al diagnóstico. 2) Caracterizar las

trayectorias longitudinales de utilización farmacológica registrada e identificar los factores asociados a su variabilidad.

9. Métodos de investigación (descripción de los métodos de investigación):

a) *Diseño del estudio: elementos clave del diseño del estudio y justificación de tal opción.*

Se llevará a cabo un estudio de cohortes longitudinal retrospectivo basado en una base de datos nacional (BIFAP). La cohorte estará compuesta por todos los pacientes con diagnóstico incidente de ND registrado entre el 1 de enero de 2019 y el 31 de diciembre de 2021, incluyendo únicamente los nuevos diagnósticos realizados durante dicho periodo. Para cada paciente, la fecha del primer diagnóstico registrado se considerará la fecha de inicio, a partir de la cual se establecerá un seguimiento individual de tres años. De este modo, los pacientes diagnosticados en 2019 serán seguidos hasta 2022, los diagnosticados en 2020 hasta 2023 y los diagnosticados en 2021 hasta 2024. Este diseño permitirá estimar el consumo farmacológico acumulado, expresado como DDD acumulada, así como analizar la evolución temporal de la utilización farmacológica registrada durante el seguimiento. Asimismo, posibilitará identificar factores basales asociados a una mayor exposición farmacológica futura y caracterizar distintas trayectorias longitudinales de utilización de fármacos en condiciones de práctica clínica real. El desarrollo, validación y comunicación de los modelos predictivos se realizarán siguiendo las recomendaciones de la guía TRIPOD+AI, y la comunicación de los aspectos observacionales generales del estudio seguirá también las recomendaciones STROBE.

b) *Entorno: población del estudio (personas, lugar, período) y criterios de selección, con justificación de los criterios de inclusión y exclusión. Si se procede a muestrear una población fuente, descripción de esta y de los métodos de muestreo;*

La población de estudio estará constituida por pacientes adultos incluidos en la base de datos BIFAP con un diagnóstico incidente de ND registrado entre el 1 de enero de 2019 y el 31 de diciembre de 2021. El estudio se desarrollará en España, en el ámbito de la práctica clínica real, utilizando la información clínica longitudinal registrada en BIFAP. La fecha de inicio se definirá como la fecha del primer registro de ND dentro del periodo de inclusión. Se incluirán sujetos de 18 o más años en la fecha de inicio, con información basal suficiente para definir las variables del estudio y con seguimiento disponible en BIFAP durante el periodo establecido. Se excluirán los menores de 18 años, los pacientes con enfermedad oncológica activa, mujeres embarazadas durante el seguimiento, los

pacientes con trastornos psiquiátricos activos previamente definidos mediante las codificaciones diagnósticas correspondientes y aquellos con información insuficiente para la definición de variables esenciales o con seguimiento incompleto que impida una observación longitudinal adecuada. No se realizará muestreo, sino que se incluirán todos los pacientes elegibles disponibles en la base de datos que cumplan los criterios de selección.

c) *Variables: de exposición, de efecto o resultado (primario y secundarios), y otras variables.*

La variable principal de resultado será el consumo farmacológico acumulado a tres años, expresado como DDD acumulada, de los fármacos utilizados en el tratamiento del dolor neuropático pertenecientes a las siguientes familias farmacológicas de interés: antidepresivos tricíclicos, inhibidores de la recaptación de serotonina y noradrenalina y ligandos de la subunidad $\alpha 2\delta$. Esta variable se calculará a partir de la información de prescripción registrada en BIFAP y se estimará tanto de forma global como por familia farmacológica. En consecuencia, la variable de resultado se interpretará como una medida de utilización farmacológica registrada y no como una medida directa de consumo real o adherencia efectiva.

De forma complementaria, se analizarán trayectorias longitudinales de utilización farmacológica registrada durante los tres años posteriores a la fecha de inicio, incluyendo, cuando la calidad y disponibilidad de los datos lo permitan, inicio de tratamiento, persistencia, discontinuación, escalada terapéutica y cambio entre familias farmacológicas.

Las variables predictoras incluirán exclusivamente características basales disponibles en BIFAP, definidas en la fecha de inicio o con anterioridad a ella, con el fin de evitar sesgos de anticipación. Entre ellas se considerarán variables demográficas, clínicas, analíticas, polimedicación y de utilización sanitaria previa al inicio de la cohorte, incluyendo edad, sexo, índice de masa corporal, HbA1c, colesterol, exposición farmacológica previa, polimedicación y frecuencia de utilización de servicios sanitarios (las exposiciones farmacológicas previas y las variables de utilización sanitaria se definirán, con carácter general, en una ventana basal de 12 meses previos a la fecha de inicio, salvo que la naturaleza de la variable aconseje otra definición operativa). La selección final de variables se limitará a aquellas con suficiente disponibilidad, calidad de registro y plausibilidad clínica.

La definición operativa de la ND, de las covariables clínicas y de las exposiciones farmacológicas se establecerá utilizando los sistemas de codificación disponibles en BIFAP, incluyendo CIE-10, SNOMED-CT y ATC/DDD, según proceda. Estas definiciones serán revisadas previamente con criterio clínico y

farmacoepidemiológico para minimizar errores de clasificación. Las definiciones finales y los códigos empleados se recogerán de forma armonizada en un anexo técnico del protocolo.

En relación con las exposiciones farmacológicas de interés, el estudio se centrará exclusivamente en los siguientes grupos terapéuticos: antidepresivos tricíclicos (principalmente amitriptilina), inhibidores de la recaptación de serotonina y noradrenalina (principalmente duloxetina y venlafaxina) y ligandos de la subunidad $\alpha 2\delta$ (pregabalina y gabapentina). No formarán parte del outcome principal otros grupos farmacológicos, como opioides, analgésicos no opioides, anestésicos locales, capsaicina u otros tratamientos analgésicos generales.

La validez de las definiciones operativas de exposición y desenlace se reforzará mediante la revisión clínica y farmacoepidemiológica previa de los códigos utilizados, la evaluación de coherencia temporal y clínica de los registros y la realización de análisis de sensibilidad frente a posibles errores de clasificación.

d) *Fuentes de datos: estrategias y fuentes de datos para determinar las exposiciones, los efectos y todas las demás variables pertinentes para los objetivos del estudio. Si el estudio utiliza fuentes de datos secundarias existentes, como archivos electrónicos de salud, o registros de pacientes, indicar toda información pertinente sobre la validez del registro y la codificación de los datos.*

Se emplearán los datos secundarios recogidos en BIFAP de las comunidades autónomas incluidas en la base. La fuente de información será principalmente la historia clínica electrónica de Atención Primaria, utilizada para la identificación de exposiciones farmacológicas, covariables clínicas y seguimiento longitudinal. Cuando estén disponibles, se incorporarán diagnósticos u otra información clínica complementaria procedente de alta hospitalaria para mejorar la identificación de eventos y covariables relevantes. La validez de las definiciones operativas se abordará mediante la revisión de códigos, la coherencia clínica de las variables seleccionadas y la evaluación de posibles errores de clasificación.

e) *Tamaño muestral: tamaño muestral previsto, precisión deseada para las estimaciones y cálculo del tamaño muestral mínimo del estudio que permitan alcanzar los objetivos del estudio con una potencia estadística preestablecida.*

Dado que se trata de un estudio observacional de base poblacional con finalidad predictiva, no se ha realizado un cálculo formal del tamaño muestral a priori basado en potencia estadística. Se incluirán todos los pacientes elegibles disponibles en BIFAP que cumplan los criterios de selección. El tamaño

muestral final vendrá determinado por el número de sujetos con información basal y seguimiento suficientes. Esta aproximación permitirá maximizar la precisión de las estimaciones y la estabilidad de los modelos predictivos.

f) Gestión de los datos.

La gestión de los datos se realizará siguiendo procedimientos estandarizados para garantizar la integridad, trazabilidad, seguridad y reproducibilidad del estudio. La extracción de datos desde BIFAP se llevará a cabo por personal autorizado, generándose un fichero de análisis anonimizado de acuerdo con las condiciones de acceso y uso de la base de datos. Los datos se organizarán en tablas estructuradas para su análisis, en las que cada fila corresponderá a un paciente y cada columna a una variable previamente definida. Asimismo, se documentarán todos los procesos realizados sobre dichos datos, incluyendo extracción, depuración, transformación, codificación y análisis.

El acceso a los datos estará limitado al equipo investigador autorizado. No se permitirá la extracción de información identificativa fuera del entorno de trabajo autorizado. Los resultados se exportarán únicamente en forma de datos agregados o resumidos, tales como medias, porcentajes e intervalos de confianza, sin compartirse en ningún caso registros individuales de pacientes.

Para la gestión analítica del estudio, el Investigador Principal elaborará un fichero de trabajo en el que constarán exclusivamente las variables necesarias para el análisis y, en su caso, un identificador interno del registro que permita la trazabilidad técnica dentro del proyecto, sin incorporar datos directamente identificativos. Este fichero se almacenará en el entorno institucional seguro correspondiente al correo corporativo del IP en CSEU La Salle, Universidad Autónoma de Madrid. Además, contará con medidas adicionales de protección, incluyendo acceso restringido y protección mediante contraseña. Se documentarán igualmente las definiciones de variables, las decisiones de depuración y recodificación y las distintas versiones de los ficheros y scripts empleados.

Control de calidad y minimización de sesgos.

El control de calidad de los datos se aplicará en todas las fases del estudio para asegurar la fiabilidad de los datos y de los resultados obtenidos. En este sentido, se llevarán a cabo varios procesos, entre los que se incluyen la definición de la cohorte, la revisión de las variables, el control de las terminologías de medicamentos mediante clasificación ATC, la revisión del código de análisis, la comprobación de la reproducibilidad y el registro de las decisiones metodológicas adoptadas.

Con el fin de reforzar la transparencia del estudio y minimizar sesgos derivados de decisiones analíticas a posteriori, incluido el riesgo de *HARKing* (formulación o reajuste de hipótesis tras conocer los resultados), una vez cuente con la aprobación por parte del comité de ética pertinente, se registrará un protocolo detallado en *Catalogue HMA-EMA of RWD studies* siguiendo las recomendaciones de *ISPE/ISPOR task force*. Asimismo, para evitar el sesgo de anticipación o *look-ahead bias*, la construcción de los modelos predictivos se basará exclusivamente en la información disponible en la fecha de inicio, sin incorporar datos posteriores al diagnóstico. Se prestará especial atención a posibles sesgos de clasificación derivados de la definición de la cohorte, de las variables clínicas y de la exposición farmacológica registrada, así como al impacto de datos incompletos o de la variabilidad del registro clínico habitual. Además, para asegurar la replicabilidad y robustez de los resultados, el Investigador Principal realizará el análisis en Python y JMC lo replicará de forma independiente en R. En caso de concordancia entre ambos análisis, se considerará que los hallazgos presentan una mayor robustez al no depender de un único entorno de programación. Por último, se realizarán análisis de sensibilidad para valorar el impacto de distintas decisiones analíticas y de posibles fuentes de variabilidad sobre los modelos desarrollados, con el fin de asegurar la estabilidad de los resultados y la calidad metodológica del informe final.

i) *Plan de análisis estadístico*

El análisis estadístico se realizará en R y Python mediante flujos de trabajo reproducibles previamente definidos. El objetivo principal será desarrollar y evaluar modelos predictivos para estimar el consumo farmacológico acumulado a tres años, expresado como DDD acumulada, de antidepresivos tricíclicos, inhibidores de la recaptación de serotonina y noradrenalina y ligandos de la subunidad $\alpha 2\delta$ en pacientes con ND. Todas las variables predictoras se definirán en la fecha de inicio o con anterioridad a ella, con el fin de evitar la incorporación de información posterior al diagnóstico y minimizar el sesgo de anticipación (look-ahead bias).

En primer lugar, se realizará un análisis descriptivo de la cohorte. Las variables cuantitativas se resumirán mediante media y desviación estándar cuando presenten distribución aproximadamente simétrica, o mediante mediana y rango intercuartílico cuando muestren distribuciones asimétricas. Las variables cualitativas se describirán mediante frecuencias absolutas y porcentajes. Asimismo, se explorará la distribución de todas las variables predictoras y de la variable principal de resultado, prestando especial atención a la asimetría, curtosis, valores extremos, inflación de ceros y problemas de escala. En la variable de desenlace se examinará específicamente la proporción de sujetos con DDD

acumulada igual a cero, así como la dispersión de los valores positivos, dado que se anticipa una distribución no normal y sesgada a la derecha.

De forma complementaria, se realizará un análisis bivariante exploratorio de asociación entre variables. La relación entre variables cuantitativas se estudiará mediante coeficientes de correlación de Pearson o Spearman, según la distribución y naturaleza de los datos. En las asociaciones entre variables categóricas y cuantitativas se utilizarán los procedimientos comparativos apropiados, y para variables categóricas se emplearán tablas de contingencia y medidas de asociación correspondientes. Este análisis tendrá una finalidad descriptiva y exploratoria, y permitirá identificar relaciones relevantes entre predictores, así como patrones iniciales de asociación con la variable de desenlace.

Además, antes de la modelización se evaluará específicamente la dependencia y redundancia entre predictores mediante matrices de correlación y revisión de pares de variables altamente correlacionadas. Esta información podrá utilizarse para simplificar el conjunto de predictores, agrupar variables redundantes o priorizar aquellas con mayor plausibilidad clínica, mejor calidad de registro y mayor utilidad predictiva.

Se evaluará también de forma sistemática el patrón de datos faltantes en todas las variables candidatas. Se cuantificará la proporción de valores ausentes por variable y por sujeto, y se explorará si su distribución sugiere ausencia completamente al azar, al azar o no al azar. Antes de la modelización se definirá un criterio de exclusión para variables con ausencia excesiva de datos o escasa utilidad clínica o analítica. Las variables finalmente seleccionadas se integrarán en un proceso de preprocesamiento reproducible, que incluirá imputación de valores faltantes, codificación de variables categóricas y transformación o escalado de variables numéricas cuando resulte necesario.

La imputación de datos faltantes se realizará exclusivamente con la información disponible en los conjuntos de entrenamiento y, en su caso, validación, sin utilizar información del conjunto de prueba. Para variables numéricas se emplearán métodos apropiados según la complejidad del modelo, como imputación por mediana o métodos más flexibles basados en vecinos próximos o algoritmos iterativos, siempre que resulten compatibles con la estabilidad del procedimiento. Para variables categóricas se utilizará imputación por la categoría más frecuente o categorías explícitas de “desconocido” cuando tenga sentido clínico y analítico. Todas estas operaciones se integrarán en pipelines cerrados para evitar fugas de información (data leakage).

Los datos se dividirán en conjuntos de entrenamiento, validación y prueba, manteniéndose el conjunto de prueba completamente independiente para la evaluación final del rendimiento

predictivo. La partición se realizará preservando, cuando resulte pertinente, la distribución de la variable de desenlace, especialmente en relación con la presencia o ausencia de exposición farmacológica acumulada. El ajuste de hiperparámetros y la selección de modelos se llevarán a cabo utilizando únicamente los conjuntos de entrenamiento y validación, o mediante validación cruzada interna, según la arquitectura analítica empleada.

Se desarrollarán y compararán dos arquitecturas predictivas preespecificadas. El modelo de referencia consistirá en un algoritmo de Extreme Gradient Boosting (XGBoost) o, en su formulación equivalente para desenlaces continuos con asimetría y exceso de ceros, un modelo de gradient boosting con función de pérdida Tweedie. Este enfoque resulta adecuado cuando la variable de resultado presenta una distribución continua muy asimétrica, con abundancia de valores nulos o bajos y cola derecha prolongada. El modelo se entrenará inicialmente de forma independiente para cada familia farmacológica de interés y, cuando proceda, también sobre la exposición acumulada global. Se optimizarán hiperparámetros como el número de árboles, la profundidad máxima, la tasa de aprendizaje, el subsampling, la regularización y el número mínimo de observaciones por nodo terminal, con el fin de equilibrar capacidad predictiva y control del sobreajuste.

Como modelo alternativo se desarrollará una red neuronal multitarea con estructura hurdle. Esta arquitectura modelizará por separado la probabilidad de presentar exposición farmacológica distinta de cero y la magnitud de la exposición acumulada entre los sujetos con valores positivos. La red incluirá un tronco compartido para aprender representaciones comunes de las variables basales y cabezas específicas para cada componente del modelo y, en su caso, para cada familia farmacológica. Este enfoque permitirá captar relaciones no lineales complejas e interacciones entre predictores, incorporando además de forma explícita la estructura inflada en ceros del desenlace. El entrenamiento se realizará con técnicas estándar de regularización, incluyendo parada temprana (early stopping), penalización y, cuando proceda, capas de dropout.

La selección del modelo final no se basará únicamente en el rendimiento predictivo bruto, sino también en criterios de estabilidad, interpretabilidad, plausibilidad clínica y robustez. Si durante el desarrollo del análisis se detectaran problemas de inestabilidad, escasez de eventos positivos o baja frecuencia de determinadas exposiciones, se considerará el uso de estrategias más parsimoniosas, incluyendo modelos lineales generalizados, modelos de dos partes más simples o aproximaciones de regresión regularizada. Cualquier adaptación del plan analítico se documentará de forma explícita.

La evaluación del rendimiento se realizará exclusivamente en el conjunto de prueba independiente. Como métricas principales para los desenlaces continuos se utilizarán el error absoluto medio (MAE)

y la raíz del error cuadrático medio (RMSE). De forma adicional, se calcularán otras métricas de apoyo, como el coeficiente de determinación (R^2) y, cuando resulte interpretable, el error porcentual absoluto medio. En el componente binario de los modelos hurdle se evaluarán métricas de discriminación y clasificación como el área bajo la curva ROC (AUC-ROC), el área bajo la curva de precisión-recall, la sensibilidad, la especificidad y, cuando proceda, medidas de calibración. La comparación entre modelos se realizará mediante remuestreo bootstrap sobre el conjunto de prueba, con estimación de intervalos de confianza del 95% para las diferencias en las métricas principales.

La interpretabilidad de los modelos se abordará de forma prioritaria. Como método principal se emplearán los valores SHAP (Shapley Additive Explanations), que permiten cuantificar la contribución individual de cada predictor a la predicción realizada por el modelo, tanto a nivel global como individual. Se obtendrán resúmenes globales de importancia de variables, gráficos de dependencia para explorar relaciones no lineales y, cuando resulte de interés, explicaciones locales de predicciones concretas. En los modelos basados en árboles, los valores SHAP proporcionarán una descomposición interpretable y consistente de la predicción; en las redes neuronales se utilizarán aproximaciones compatibles, como DeepSHAP, siempre que su comportamiento resulte estable. De forma complementaria, se emplearán otros métodos de explicación, como la importancia por permutación, y, cuando sea útil, gráficos de dependencia parcial o curvas de efecto acumulado local, con el fin de combinar explicabilidad global, local y basada en sensibilidad del modelo.

Se realizarán además análisis de sensibilidad para valorar la robustez de los hallazgos. Entre ellos se contemplarán comparaciones entre distintos métodos de imputación, análisis con diferentes definiciones operativas de la exposición farmacológica, exclusión de variables con alto grado de ausencia o baja fiabilidad y evaluación del efecto de simplificar o agrupar determinadas familias farmacológicas. También se explorará la estabilidad del rendimiento al modificar la estrategia de partición de los datos o el conjunto final de predictores. Asimismo, cuando el tamaño muestral efectivo lo permita, se realizarán análisis estratificados por subgrupos clínicamente relevantes y se explorará la posible heterogeneidad del rendimiento predictivo y de la relevancia de los predictores entre dichos subgrupos.

Finalmente, todos los análisis se documentarán de forma exhaustiva, incluyendo decisiones sobre selección de variables, preprocesamiento, ajuste de hiperparámetros, criterios de evaluación y modificaciones justificadas respecto al plan inicial.

h) Limitaciones de los métodos de investigación.

El presente estudio presenta varias limitaciones inherentes a su diseño y a la fuente de datos utilizada. En primer lugar, al tratarse de un estudio observacional retrospectivo basado en registros clínicos rutinarios, la calidad de la información dependerá de la exactitud, exhaustividad y consistencia del registro en BIFAP. En este sentido, pueden existir errores de codificación, variabilidad entre profesionales o centros, así como diferencias en la completitud de determinadas variables clínicas y analíticas.

En segundo lugar, el consumo farmacológico se estimará a partir de la prescripción registrada y de su correspondiente equivalencia en DDD, por lo que no será posible confirmar la adherencia real ni la toma efectiva del tratamiento por parte del paciente. En consecuencia, la variable principal debe interpretarse como una medida de utilización farmacológica registrada y no como una medida directa de consumo real. Esta limitación es especialmente relevante en tratamientos crónicos, en los que pueden existir interrupciones, baja adherencia o cambios terapéuticos no completamente reflejados en la toma efectiva del medicamento.

En tercer lugar, aunque el análisis se restringirá a antidepresivos tricíclicos, inhibidores de la recaptación de serotonina y noradrenalina y ligandos de la subunidad $\alpha 2\delta$, estos grupos farmacológicos pueden utilizarse en algunos casos por indicaciones distintas al dolor neuropático. Por tanto, existe la posibilidad de confusión por indicación y de cierta clasificación errónea de la exposición, especialmente en ausencia de una confirmación directa del motivo clínico de prescripción en todos los casos. Esta limitación se abordará mediante una definición rigurosa de la cohorte, la selección clínica de covariables relevantes y análisis de sensibilidad cuando resulte posible.

Asimismo, pueden persistir confusores residuales derivados de variables no disponibles o registradas de forma incompleta en la base de datos, como la intensidad del dolor, su duración, determinadas variables funcionales, factores psicosociales, nivel socioeconómico, gravedad clínica de la neuropatía o preferencias del profesional prescriptor. Aunque se incorporarán múltiples covariables basales para mejorar el ajuste de los modelos, no puede excluirse completamente la influencia de factores no medidos.

Otra limitación relevante es la posible presencia de datos faltantes en algunas variables predictoras. Aunque estos se abordarán mediante estrategias de imputación y análisis de sensibilidad, la ausencia de información puede introducir incertidumbre adicional y afectar a la estabilidad de algunos modelos, especialmente en subgrupos con menor tamaño muestral efectivo.

Desde el punto de vista analítico, los modelos de aprendizaje automático permiten captar relaciones complejas y no lineales entre variables, pero también pueden ser más sensibles al sobreajuste, a la distribución de los datos y a la baja frecuencia de determinadas exposiciones. Para minimizar este riesgo, se utilizarán procedimientos de partición entrenamiento-validación-prueba, ajuste interno de hiperparámetros, control del data leakage, validación interna y análisis de sensibilidad. Asimismo, con el fin de reforzar la transparencia del proceso analítico y reducir el riesgo de decisiones post hoc, incluido el HARKing (formulación o reajuste de hipótesis tras conocer los resultados), el protocolo y el plan analítico se registrarán previamente en *Catalogue HMA-EMA of RWD studies* siguiendo las recomendaciones de *ISPE/ISPOR task force*. No obstante, el rendimiento observado puede depender en parte de las características específicas de la cohorte analizada y requerirá una interpretación prudente.

Finalmente, debido al carácter observacional y retrospectivo del estudio, los resultados permitirán identificar asociaciones y patrones predictivos, pero no establecer relaciones causales. En consecuencia, los hallazgos deberán interpretarse como evidencia orientada a la predicción y a la generación de conocimiento clínicamente útil en práctica real, y no como prueba de efecto causal de las variables estudiadas sobre la exposición farmacológica futura.

10. Protección de las personas sometidas al estudio: salvaguardias que permitan cumplir los requisitos nacionales y de la Unión Europea para garantizar el bienestar y los derechos de los sujetos participantes en estudios observacionales con medicamentos:

a) Evaluación beneficio-riesgo para los sujetos de investigación, en su caso.

Dado que se trata de un estudio observacional retrospectivo basado en el análisis de datos ya registrados en BIFAP, no se realizará ninguna intervención sobre los pacientes ni se modificará su seguimiento clínico habitual. En consecuencia, no se prevén riesgos directos para los sujetos derivados de su participación en el estudio. El principal riesgo potencial se relaciona con la confidencialidad y el tratamiento de la información, si bien este riesgo se considera mínimo al trabajarse con datos anonimizados y bajo condiciones de acceso controlado. El beneficio esperado es de carácter indirecto y colectivo, al contribuir a mejorar el conocimiento sobre la exposición farmacológica en ND y sobre sus determinantes, lo que podría favorecer en el futuro una toma de decisiones clínicas más informada y una utilización más segura y eficiente de los tratamientos.

b) Consideraciones sobre información a los sujetos y consentimiento informado.

Dada la naturaleza retrospectiva del estudio y la procedencia secundaria de los datos a partir de BIFAP, no se prevé la obtención de consentimiento informado individual de los pacientes. Los datos utilizados en el estudio se facilitarán en el marco de las condiciones de acceso de BIFAP y no permitirán la identificación directa de los sujetos por parte del equipo investigador. El proyecto se someterá a los procedimientos de evaluación, autorización y supervisión que resulten aplicables de acuerdo con la normativa vigente y con las condiciones de acceso a la fuente de datos.

c) Confidencialidad de los datos: las condiciones de acceso y tratamiento de datos de carácter personal, incluyendo, en el caso de transmisión de los datos personales de pacientes españoles a un tercer Estado, acreditación del cumplimiento de la normativa europea sobre protección de datos de carácter personal. Cuando se trate de una investigación con datos anónimos o que se han sometido a un tratamiento de seudonimización se establecerá en el protocolo del estudio el procedimiento previsto para ello;

El tratamiento de los datos se realizará conforme a las condiciones de acceso y uso establecidas para BIFAP y de acuerdo con la normativa vigente en materia de protección de datos personales. El equipo investigador trabajará exclusivamente con un fichero de análisis anonimizado o sin identificadores directos, sin posibilidad de reidentificación de los pacientes en el contexto del estudio. El acceso a la información estará restringido al personal investigador autorizado y los resultados se presentarán únicamente de forma agregada, evitando en todo momento la difusión de información individual. Los datos no se cederán a terceros ni se transferirán fuera del marco autorizado para el proyecto. Asimismo, se adoptarán las medidas técnicas y organizativas necesarias para garantizar la seguridad, integridad y confidencialidad de la información durante todas las fases del estudio.

d) Interferencia con los hábitos de prescripción del médico:

El estudio se limitará exclusivamente al análisis secundario de datos ya existentes en BIFAP. Por tanto, no implicará modificación alguna de la prescripción farmacológica, ni interferirá en la práctica clínica habitual, ni influirá en la toma de decisiones diagnósticas o terapéuticas de los profesionales sanitarios.

11. Gestión y notificación de reacciones adversas y demás eventos relevantes que aparezcan durante el estudio: *el protocolo deberá incluir la especificación de los casos de sospechas de reacciones*

adversas a los medicamentos objeto del estudio que, en su caso, deben ser registrados sistemáticamente por el profesional sanitario y por tanto transmitidos al promotor. Deberá asimismo recordar la importancia de comunicar el resto de sospechas de reacciones adversas que detecte el profesional en relación con cualquier medicamento en el transcurso del estudio, y de las vías para que realice su comunicación al Sistema Español de Farmacovigilancia.

Dado que se trata de un estudio observacional retrospectivo basado en datos previamente registrados en historias clínicas electrónicas y que no implica ninguna intervención sobre los pacientes ni contacto directo con ellos, la identificación prospectiva, el seguimiento activo y la notificación sistemática de sospechas de reacciones adversas no constituyen una finalidad específica del estudio. Asimismo, el estudio no modificará la prescripción, dispensación ni administración de los medicamentos objeto de análisis, ni interferirá en la práctica clínica habitual.

No obstante, si durante el proceso de revisión, depuración o análisis de los datos se identificaran hallazgos relevantes relacionados con la seguridad de los medicamentos que, por su naturaleza, frecuencia, gravedad o falta de concordancia con la información previamente conocida, pudieran resultar de interés desde el punto de vista farmacovigilante, dichos hallazgos serán valorados por el equipo investigador y gestionados de acuerdo con los procedimientos y cauces que resulten aplicables en el marco del estudio y de la normativa vigente.

En cualquier caso, este estudio no sustituye ni modifica las obligaciones habituales de notificación de sospechas de reacciones adversas por parte de los profesionales sanitarios en su práctica clínica, que deberán seguir realizándose a través de las vías establecidas en el Sistema Español de Farmacovigilancia.

12. Plan de trabajo (tareas, hitos y cronología del estudio).

Etapas / Hito	Actividades Principales	Cronograma
a) Inicio de la recogida de datos	Aprobación BIFAP, envío a Comité de Ética, autorizaciones de acceso y definición final de variables.	Abril – Julio 2026
b) Final de la recogida de datos	Extracción de base de datos, limpieza, validación y estructuración del dataset analítico.	Agosto – Septiembre 2026

c) Informes de situación	Análisis descriptivo inicial, caracterización de la cohorte y evaluación de datos faltantes.	Noviembre – Diciembre 2026
d) Informes intermedios	Modelización predictiva (ML/Regresión), análisis de tendencias y análisis de sensibilidad.	Diciembre 2026 – Enero 2027
e) Informe final de resultados	Interpretación de resultados, redacción del informe final y envío del manuscrito científico.	Febrero 2027

Asignación de roles y responsabilidades del equipo investigador:

La ejecución del proyecto se articulará mediante una estructura jerárquica y colaborativa destinada a garantizar el rigor científico, la calidad metodológica y la reproducibilidad de los resultados. MMA, como Investigador Principal y experto en ciencia de datos aplicada a ciencias de la salud, asumirá la dirección científica y técnica del estudio, la coordinación del equipo, la supervisión del dataset procedente de BIFAP, la definición de la estrategia analítica, la ejecución del análisis principal en Python y la redacción final del informe. JMC, experto en bioestadística y metodología cuantitativa, será responsable de la supervisión metodológica y de la replicación independiente de los análisis en R, así como de la revisión del cumplimiento de las guías STROBE y TRIPOD+AI. STB colaborará en la redacción inicial del protocolo, la tramitación ética y la solicitud de acceso a BIFAP, además de participar en la estadística descriptiva, la documentación analítica y el borrador del informe final. LMG, como experto clínico en dolor y neuropatías, participará en la revisión de la calidad de los datos, la interpretación de los resultados, la discusión de los hallazgos y la revisión crítica de los documentos científicos derivados del proyecto.

13. Planes de difusión y comunicación de los resultados del estudio.

Los resultados del estudio se difundirán de forma agregada, garantizando en todo momento la confidencialidad de la información y el cumplimiento de la normativa aplicable en materia de protección de datos. En ningún caso se comunicarán datos individuales ni información que permita la identificación directa o indirecta de los pacientes incluidos en la base de datos.

La estrategia de difusión incluirá, en primer lugar, la elaboración de un informe final de resultados, que recogerá los principales hallazgos del estudio, sus implicaciones clínicas y metodológicas, así como sus limitaciones. Asimismo, se prevé la preparación de al menos un manuscrito científico para su envío

a una revista biomédica revisada por pares, preferentemente en el ámbito de la epidemiología clínica, la farmacología, la diabetología o el dolor neuropático.

Además, los resultados podrán presentarse en congresos y reuniones científicas nacionales e internacionales relacionadas con la diabetes, el dolor neuropático, la farmacología clínica, la epidemiología y la inteligencia artificial aplicada a la salud. La comunicación científica seguirá criterios de transparencia y calidad metodológica, incluyendo las recomendaciones de las guías STROBE y TRIPOD+AI, según corresponda al tipo de resultado presentado. Asimismo, se comunicará el cierre del estudio y sus resultados a los órganos o plataformas que resulten aplicables conforme a las condiciones de autorización del proyecto y a la normativa vigente. En particular, una vez finalizado el estudio, el Investigador Principal se compromete a comunicar la publicación de los resultados obtenidos a BIFAP a través de los cauces establecidos. Del mismo modo, si la información resultante pudiera suponer una modificación de la relación beneficio-riesgo de un medicamento, dichos resultados se comunicarán a la AEMPS con carácter previo e independiente de su eventual publicación científica o registro.

Con el fin de reforzar la transparencia y la reproducibilidad, se promoverá la difusión de los aspectos metodológicos generales del estudio y de sus principales decisiones analíticas a través de plataformas de ciencia abierta, siempre dentro de los límites permitidos por las condiciones de uso de BIFAP y por la normativa vigente. En todo caso, cualquier difusión se realizará exclusivamente sobre resultados agregados y nunca sobre datos individuales.

Finalmente, se contempla la posibilidad de comunicar los hallazgos más relevantes a los agentes académicos y clínicos interesados, con el objetivo de favorecer la transferencia de conocimiento y contribuir a una mejor comprensión de los patrones de utilización farmacológica en ND en condiciones de práctica clínica real. Asimismo, el Investigador Principal, como responsable último de la integridad de la investigación desde su inicio hasta su publicación, velará por el cumplimiento de las recomendaciones del *International Committee of Medical Journal Editors* en materia de autoría de las comunicaciones científicas y por la declaración de los potenciales conflictos de interés de todos los autores en relación con los medicamentos objeto del estudio.

14. Referencias.

1. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *New England Journal of Medicine*. 2019 Apr 4;380(14):1347–58. doi:10.1056/nejmra1814259 PubMed PMID: 30943338.
2. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA*. 2018 Apr 3;319(13):1317–8. doi:10.1001/jama.2017.18391 PubMed PMID: 29532063.
3. Deo RC. Machine Learning in Medicine. *Circulation*. 2015 Nov 17;132(20):1920. doi:10.1161/CIRCULATIONAHA.115.001593 PubMed PMID: 26572668.
4. Anderson SG, Narayanan RP, Malipatil NS, Roberts H, Dunn G, Heald AH. Socioeconomic deprivation independently predicts painful diabetic neuropathy in type 2 diabetes. *Exp Clin Endocrinol Diabetes*. 2015 Apr 22;123(7):423–7. doi:10.1055/s-0035-1549966 PubMed PMID: 26069072.
5. Anderson SG, Malipatil NS, Roberts H, Dunn G, Heald AH. Socioeconomic deprivation independently predicts symptomatic painful diabetic neuropathy in type 1 diabetes. *Prim Care Diabetes*. 2014;8(1):65–9. doi:10.1016/j.pcd.2013.08.004 PubMed PMID: 24211151.
6. Hébert HL, Veluchamy A, Torrance N, Smith BH. Risk factors for neuropathic pain in diabetes mellitus. *Pain*. 2017 Apr 1;158(4):560–8. doi:10.1097/j.pain.0000000000000785 PubMed PMID: 27941499.
7. Vollert J, Kleykamp BA, Farrar JT, Gilron I, Hohenschurz-Schmidt D, Kerns RD, et al. Real-world data and evidence in pain research: a qualitative systematic review of methods in current practice. *Pain Rep*. 2023 Mar 1;8(2):e1057. doi:10.1097/PR9.0000000000001057 PubMed PMID: 36741790.
8. Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L. Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. *BMJ Open Diabetes Res Care*. 2022 Sep 16;10(5):2946. doi:10.1136/BMJDRC-2022-002946 PubMed PMID: 10.1136/bmjdr-2022-002946.
9. Pai YW, Tang CL, Lin CH, Lin SY, Lee IT, Chang MH. Glycaemic control for painful diabetic peripheral neuropathy is more than fasting plasma glucose and glycated haemoglobin. *Diabetes Metab*. 2021 Feb 1;47(1):101158. doi:10.1016/J.DIABET.2020.04.004 PubMed PMID: 32413414.

10. Cavero-Redondo I, Peleteiro B, Álvarez-Bueno C, Rodríguez-Artalejo F, Martínez-Vizcaíno V. Glycated haemoglobin A1c as a risk factor of cardiovascular outcomes and all-cause mortality in diabetic and non-diabetic populations: a systematic review and meta-analysis. *BMJ Open*. 2017 Jul 1;7(7):e015949. doi:10.1136/BMJOPEN-2017-015949 PubMed PMID: 28760792.
11. Sloan G, Selvarajah D, Tesfaye S. Pathogenesis, diagnosis and clinical management of diabetic sensorimotor peripheral neuropathy. *Nature Reviews Endocrinology* 2021 17:7. 2021 May 28;17(7):400–20. doi:10.1038/s41574-021-00496-z PubMed PMID: 34050323.
12. Vollert J, Spivack J, Adamson B, Baron R, Farrar JT, Gilron I, et al. Real-world data and evidence in pain research: an IMMPACT comprehensive review and best practice recommendations. *Pain*. 2025;167(1):8. doi:10.1097/J.PAIN.0000000000003798 PubMed PMID: 40893022.
13. Lai YR, Chiu WC, Cheng BC, Yu IH, Lin TY, Chiang HC, et al. Impact of HbA1c variability and time-in-range fluctuations on large and small nerve fiber dysfunction in well-controlled type 2 diabetes: A prospective cohort observational study. *J Diabetes Investig*. 2025 Aug 1;16(8):1507. doi:10.1111/JDI.70079 PubMed PMID: 40396765.
14. Aburayyan W, Zakaraya Z, Hamad M, Majali IS, Dayyih WA, Seder N, et al. Improving HbA1c Levels by Methylcobalamin Vitamin in Diabetic Volunteers, Combined with Dapagliflozin as Type 2 Diabetes Mellitus Routine Treatment: A Controlled Randomized, Double-blind Trial. *Iran J Med Sci*. 2025 May 1;50(5):324. doi:10.30476/IJMS.2024.101606.3423 PubMed PMID: 40433181.
15. Strom A, Strassburger K, Ziegler D, Sipola G, Prystupa K, Wagner R, et al. Changes Over 10 Years in Peripheral Nerve Function in People With Well-Controlled Type 2 Diabetes and Those With Normal Glucose Tolerance. *Neurology*. 2025 May 29;104(12):e213780. doi:10.1212/WNL.0000000000213780 PubMed PMID: 40440593.
16. Kec D, Rajdova A, Raputova J, Adamova B, Srotova I, Nekvapilova EK, et al. Risk factors for depression and anxiety in painful and painless diabetic polyneuropathy: A multicentre observational cross-sectional study. *Eur J Pain*. 2021 Feb 1;26(2):370. doi:10.1002/EJP.1865 PubMed PMID: 34592017.
17. Truini A, Spallone V, Morganti R, Tamburin S, Zanette G, Schenone A, et al. A cross-sectional study investigating frequency and features of definitely diagnosed diabetic painful

- polyneuropathy. *Pain*. 2018;159(12):2658–66. doi:10.1097/J.PAIN.0000000000001378 PubMed PMID: 30161042.
18. Abbott CA, Malik RA, Van Ross ERE, Kulkarni J, Boulton AJM. Prevalence and Characteristics of Painful Diabetic Neuropathy in a Large Community-Based Diabetic Population in the U.K. *Diabetes Care*. 2011 Oct;34(10):2220. doi:10.2337/DC11-1108 PubMed PMID: 21852677.
 19. Koponen ME, Senaratne DNS, Hébert HL, Smith BH, Colvin LA. Identifying predictors of neuropathic pain medication prescribing, adherence, and discontinuation: a systematic review and meta-analysis. *Br J Anaesth*. 2025 Nov;0(0). doi:10.1016/J.BJA.2025.10.047 PubMed PMID: 41314943.
 20. Savelieff MG, Elafros MA, Viswanathan V, Jensen TS, Bennett DL, Feldman EL. The global and regional burden of diabetic peripheral neuropathy. *Nat Rev Neurol*. 2025 Jan 1;21(1):17–31. doi:10.1038/S41582-024-01041-Y PubMed PMID: 39639140.
 21. Yang Y, Zhao B, Wang Y, Lan H, Liu X, Hu Y, et al. Diabetic neuropathy: cutting-edge research and future directions. *Signal Transduction and Targeted Therapy* 2025 10:1. 2025 Apr 25;10(1):132-. doi:10.1038/s41392-025-02175-1 PubMed PMID: 40274830.
 22. Słyk S, Kamińska-Słyk N, Domitrz I. Carpal Tunnel Syndrome Diagnostics Revisited: A Comprehensive Review. *Int J Clin Pract*. 2025 Jan 1;2025(1):2893823. doi:10.1155/IJCP/2893823
 23. Hughes RAC. Peripheral neuropathy. *BMJ : British Medical Journal*. 2002 Feb 23;324(7335):466. doi:10.1136/BMJ.324.7335.466 PubMed PMID: 11859051.