



UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y
DISEÑO INDUSTRIAL

Grado en Ingeniería Electrónica Industrial y Automática

TRABAJO FIN DE GRADO

**CONEXIÓN ENTRE EL DECLIVE
COGNITIVO EN SUJETOS CON LA
ENFERMEDAD DE PARKINSON NOVO Y
LOS CONSTRUCTOS BIOLÓGICOS DE LA
ENFERMEDAD DE ALZHEIMER**

Jaime Palomino Vaquero

Tutor: Carlos Platero Dueñas

Departamento: Ingeniería Eléctrica, Electrónica, Automática y Física
Aplicada

Madrid, Junio, 2022



UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y
DISEÑO INDUSTRIAL

Grado en Ingeniería Electrónica y Automática Industrial

TRABAJO FIN DE GRADO

CONEXIÓN ENTRE EL DECLIVE
COGNITIVO EN SUJETOS CON LA
ENFERMEDAD DE PARKINSON NOVO Y
LOS CONSTRUCTOS BIOLÓGICOS DE LA
ENFERMEDAD DE ALZHEIMER

Firma Autor

Firma Cotutor (si lo hay)

Firma Tutor

Copyright ©2022. Jaime Palomino Vaquero

Esta obra está licenciada bajo la licencia Creative Commons

Atribución-NoComercial-SinDerivadas 3.0 Unported (CC BY-NC-ND 3.0). Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/3.0/deed.es> o envíe una carta a Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, EE.UU.

Todas las opiniones aquí expresadas son del autor, y no reflejan necesariamente las opiniones de la Universidad Politécnica de Madrid.

Título: Conexión entre el declive cognitivo en sujetos con la enfermedad de Parkinson novo y los constructos biológicos de la enfermedad de Alzheimer.

Autor: Jaime Palomino Vaquero

Tutor: Carlos Platero Dueñas

EL TRIBUNAL

Presidente:

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Trabajo Fin de Grado el día ... de de ... en, en la Escuela Técnica Superior de Ingeniería y Diseño Industrial de la Universidad Politécnica de Madrid, acuerda otorgarle la CALIFICACIÓN de:

VOCAL

SECRETARIO

PRESIDENTE

Agradecimientos

A mi abuela Angelita, persona con enfermedad neurodegenerativa. A ella va dedicado todo este trabajo. Por siempre motivarme a salir de mi comodidad, a arriesgarme y a conseguir todo lo que quiero. Ahora puedo saber que estás orgullosa. Lo hemos hecho juntos.

Al Jefe, por ser el motor de estos años de estudio. Por depositar en Él toda la ansiedad. Por poner a esas personas que han servido de apoyo en el camino. Por saber que así todo es más fácil.

A mis padres y hermanos. Por la confianza ciega en mí y por inculcar los valores de vida necesarios para luchar por mis metas, por el respeto a los demás y por el deber de cumplir con lo que uno se propone. Por aguantar mi carácter y por sacarme alegría en el agobio.

A mi familia. Por recordarme de dónde soy, por poner los pies en tierra. Por ser luchadores y por estar alegres con mis progresos.

A mis amigos, a los que he ganado en estos años de universidad y que se quedan para siempre. Y como no, a todos los de siempre, a los que traía conmigo antes. Por apoyarme sin condición, por empujarme hacia adelante y ser cada logro mío, orgullo para ellos.

A Raquel y Lidia, por soportarme en estos años cada día. Por verme en las caídas frustradas y en las celebraciones más eufóricas. Seguimos celebrando la vida juntos.

Y por último, gracias a mi tutor Carlos. Por su gran paciencia, su orientación, su forma de querer mostrarnos esta área. Gracias por la oportunidad de sumergirme en el mundo de la inteligencia artificial y el análisis de datos.

Gracias a todas las personas que me he cruzado en este tiempo y me ha aportado algo, a la gente que cree en mí.

Hagan todo con amor. Con todo lo que tengo.

Gracias.

Resumen

La enfermedad de Parkinson es el segundo trastorno neurodegenerativo más común. Este trastorno deteriora la vida de las personas y de todos los que se encuentran a su alrededor.

Este proyecto tiene como objetivo el análisis de los diferentes conjuntos de biomarcadores al inicio del estudio en diferentes grupos clínicos para la elaboración de modelos predictivos que diagnostiquen la enfermedad. Para ello, se llevan a cabo dos líneas de trabajo.

1. **Estudio transversal.** Análisis de los biomarcadores para la determinación de los biomarcadores y variables clínicas que más discriminen en el progreso del declive cognitivo de la PD y su conexión con los constructos biológicos de la AD.
2. **Estudio longitudinal.** Análisis de los biomarcadores mediante modelos predictivos para realizar un diagnóstico de la evolución cognitiva de la enfermedad en sujetos con PD. Con ello se ofrece un modelo temporal de la progresión del declive cognitivo en los sujetos con la PD, estimando los tiempos de conversión a MCI y demencia en los pacientes PD.

Palabras clave: Enfermedad de Parkinson, Enfermedad de Alzheimer, PPMI, ADNI, ATN, MRI, CSF, DAT, modelo predictivo, estudio longitudinal.

Abstract

Parkinson's disease is the second most common neurodegenerative disorder. This disorder impairs the lives of people and everyone around them. This project aims to analyze the different sets of biomarkers at the beginning of the study in different clinical groups to develop predictive models that diagnose the disease. To do this, two lines of work are carried out.

1. **Cross-sectional study.** Analysis of biomarkers for the determination of biomarkers and clinical variables that most discriminate in the progress of cognitive decline of PD and its connection with the biological constructs of AD.
2. **Longitudinal study.** Analysis of biomarkers using predictive models to make a diagnosis of the cognitive evolution of the disease in subjects with PD. This provides a temporal model of the progression of cognitive decline in PD subjects, estimating the conversion times to MCI and dementia in PD patients.

Keywords: Parkinson Disease, Alzheimer Disease, PPMI, ADNI, ATN, MRI, CSF, DAT, predictive models, longitudinal study,

Índice general

Agradecimientos	IX
Resumen	XI
Abstract	XIII
Índice	XVIII
1. Introducción	1
1.1. Neurodegeneración.	1
1.2. La enfermedad de Parkinson.	2
1.2.1. Síntomas.	2
1.2.2. Factores de riesgo.	3
1.2.3. Diagnóstico.	4
1.2.4. Tratamiento.	5
1.2.5. Impacto socioeconómico.	5
1.3. Motivación del proyecto.	6
1.4. Naturaleza del estudio.	6
1.5. Objetivos.	7
1.6. Estructura del documento	7
2. Estado de la técnica.	9
2.1. Parkinson Disease-Associated Cognitive Impairment [7].	9
2.1.1. Introducción	9
2.1.2. Epidemiología	10
2.1.3. Diagnóstico, cribado y prevención	11
2.2. Toward a biological definition of Alzheimer's disease [6].	12
2.2.1. Introducción	12
2.2.2. Alzheimer como cambios neuropatológicos	13
2.2.3. Definición de la Enfermedad del Alzheimer	14
2.2.4. Puesta en escena	14
2.2.5. Perfiles y categoría de biomarcadores	14
2.2.6. Características y limitaciones de los biomarcadores	15
2.2.7. CSF T-tau y P-tau	15
2.3. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study [2].	16
2.3.1. Introducción	16

2.3.2.	Métodos	16
2.3.3.	Resultados y Discusión.	17
2.4.	Brain Amyloid Contribution to Cognitive Dysfunction in Early-Stage Parkinson's Disease: The PPMI Dataset [16].	20
2.4.1.	Introducción.	20
2.4.2.	Métodos	21
2.4.3.	Características neuropsicológicas y clínicas.	21
2.4.4.	Resultados y discusión.	22
2.5.	Categorical predictive and disease progression modeling in the early stage of Alzheimer's disease [42].	25
2.5.1.	Introducción.	25
2.5.2.	Materiales.	25
2.5.3.	Métodos.	26
2.5.4.	Resultados.	28
2.6.	Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data [26].	29
2.6.1.	Introducción.	29
2.6.2.	Material y métodos.	29
2.6.3.	Resultados.	32
2.7.	Estimating long-term multivariate progression from short-term data [34]	34
2.7.1.	Introducción.	34
2.7.2.	Supuestos del modelo.	34
2.7.3.	El Algoritmo.	35
2.7.4.	Resultados y discusión.	35
2.8.	Conclusión del capítulo	36
3.	Datos empleados, herramientas y materiales	39
3.1.	Introducción.	39
3.2.	Datos empleados.	39
3.2.1.	Base de datos. PPMI.	39
3.3.	Herramientas.	40
3.3.1.	Matlab.	40
3.3.2.	Filezilla.	41
3.3.3.	PuTTY.	42
3.3.4.	Xming.	42
3.3.5.	LaTeX.	42
3.3.6.	Procesamiento en CeSViMa. Magerit-3.	43
4.	Métodos.	45
4.1.	Introducción.	45
4.2.	Determinación deterioro cognitivo leve (MCI).	45
4.3.	Tiempos de conversión y censura.	45
4.4.	Análisis al inicio del estudio.	46
4.4.1.	Análisis al inicio del estudio de sujetos PD y HC	47
4.4.2.	Análisis al inicio del estudio de sujetos PD-A+, PD-A-, HC-A+ y HC-A-	47
4.4.3.	Análisis al inicio del estudio de sujetos PD-CU, PD-MCI y PDD	48

4.4.4.	Predictores y factores de riesgo	48
4.4.5.	Perfil AT en sujetos PD	50
4.5.	Análisis estadístico.	53
4.5.1.	Realización de modelos predictivos.	54
4.5.2.	Realización de modelos de progresión. GRACE	56
5.	Resultados	61
5.1.	Introducción	61
5.2.	Resultados del análisis al inicio del estudio	61
5.2.1.	Análisis al inicio del estudio de sujetos PD y HC	61
5.2.2.	Análisis al inicio del estudio de sujetos PD-A+, PD-A-, HC-A+ y HC-A-	64
5.2.3.	Análisis al inicio del estudio de sujetos PD-CU y PD-MCI y PDD.	67
5.2.4.	Predictores y factores de riesgo	68
5.2.5.	Perfil AT en sujetos PD	69
5.3.	Resultados del análisis multivariante.	71
5.3.1.	Modelos predictivos.	72
5.3.2.	Realización de los modelos de progresión. GRACE	76
6.	Discusiones	81
6.1.	Discusiones	81
6.1.1.	Población PD-CU, PD-MCI y PDD al inicio del estudio	82
6.1.2.	Perfiles AT en sujetos PD	84
6.1.3.	Factores de riesgo	86
6.1.4.	Modelos predictivos al inicio del estudio	87
6.1.5.	Historia natural mediante los modelos de progresión. GRACE	88
6.2.	Conclusiones.	92
6.3.	Limitaciones del trabajo.	93
6.4.	Desarrollos futuros	93
A.	Análisis estadístico aplicado a ciencias de la salud	95
A.1.	Introducción	95
A.2.	Conceptos básicos	95
A.2.1.	Inferencia estadística	95
A.2.2.	Pruebas de hipótesis	96
A.2.3.	Estadísticos y p-valor	97
A.2.4.	Tratamiento de errores	98
A.2.5.	FDR y q-valores	98
A.2.6.	Calidad del experimento	98
A.3.	Modelos de análisis de la varianza	99
A.4.	Métodos no paramétricos.	101
A.4.1.	Prueba U de Mann-Whitney.	101
A.5.	Correlación lineal de Pearson.	102
A.5.1.	Interpretación del coeficiente de correlación.	103

B. Comparaciones múltiples	105
B.1. Introducción	105
B.2. Comparaciones múltiples de medias.	105
B.2.1. Comparaciones gráficas de medias	105
B.2.2. Comparaciones basadas en la distribución t	106
B.2.3. Test de rangos múltiples	108
C. Empleo de Magerit	111
C.1. Introducción	111
C.2. Registro y acceso a Magerit	111
C.3. SLURM	111
C.4. Mandatos SLURM	112
C.5. Carga de módulos	112
C.6. Parámetros de ejecución	113
C.7. Directorio SCRATCH	114
C.8. Notificación de estado y redirección de salidas	114
D. Diccionario de siglas y acrónimos	115
D.1. Siglas	115
Bibliografía	121

Índice de figuras

1.1. Factores de riesgo modificables y no modificables asociados al desarrollo de la enfermedad de Parkinson.	4
2.1. Tabla 1. Características demográficas, clínicas y neuropsicológicas de acuerdo con el estado amiloide.	23
2.2. Tabla 2. SUVR regionales de forbetaben según el estado amiloide. . .	24
2.3. Significación estadística no corregida: log10 negativo (p-valor): mapas que comparan las tasas de adelgazamiento cortical longitudinal entre sujetos HC (N = 210) y AD (N = 188) (de toda la muestra de ADNI) visualizados en la superficie pial de la plantilla FreeSurfer (fsaverage): (A) método ST-LME y (B) método X-Slope. El hemisferio izquierdo se muestra a la izquierda y el hemisferio derecho a la derecha. Los vértices que tienen un p-valor sin corregir inferior a 0.05 se muestran en color. Las filas impares muestran las vistas lateral, superior y anterior. Las filas pares muestran las vistas medial, inferior y posterior. La barra de color muestra el valor de significación correspondiente. . .	33
2.4. Regiones corticales que muestran una diferencia estadísticamente significativa en el adelgazamiento longitudinal entre sujetos HC y AD (en rojo) en toda la muestra de ADNI. Estos mapas se derivaron de umbralizar los valores que se muestran en Figura 2.3 con una corrección FDR en $q = 0,05$. (A) método ST-LME y (B) método X-Slope. ST-LME revela un grado mucho más amplio de adelgazamiento significativo en la AD.	33

- 2.5. (A) Sujetos A+ de ADNIA. Portadores de APOE ϵ 4 en ADNI. Los paneles superiores muestran cada una de las trayectorias medias superpuestas sobre las observaciones a nivel de sujeto de 579 A+ y 570 APOE ϵ 4 individuos, coloreados por diagnóstico. Los colores en el panel superior representan el diagnóstico en la línea de base de ADNI: cognitivamente normal (CN) en azul oscuro, deterioro cognitivo leve temprano (EMCI) en azul claro, deterioro cognitivo leve tardío (LMCI) en rojo claro y enfermedad de Alzheimer (AD) en oscuro rojo. Las regiones sombreadas en gris, donde están visibles en los paneles superiores, representan bandas de confianza del 95% de arranque. Los paneles centrales muestran todas las trayectorias a la vez. A la izquierda, el tiempo se ha desplazado de modo que el tiempo cero representa el tiempo en el que la trayectoria media de la suma de cajas de la escala de calificación clínica de demencia (*CDRSB, Clinical Dementia Rating Scale Sum of Boxes*) (no se muestra) alcanza el percentil 80. A la derecha, el tiempo se ha ajustado utilizando trayectorias de miniexamen del estado mental (PAQUID) a largo plazo, de modo que el tiempo cero representa el tiempo estimado hasta el inicio de la demencia. Los paneles inferiores muestran tasas de cambio estandarizadas por desviación estándar residual (SD). CSF, líquido cefalorraquídeo; A β , amiloide- β ; p-tau, tau fosforilada; PiB, compuesto B de Pittsburgh; PET, tomografía por emisión de positrones; FDG, fluorodesoxiglucosa; ADAS13, la Escala de Evaluación de la Enfermedad de Alzheimer de 13 ítems–Subescala Cognitiva; MMSE, Mini-Examen del Estado Mental; FAQ, Cuestionario de Actividades Funcionales del Estudio Cooperativo de la Enfermedad de Alzheimer; RAVLT, Test de Aprendizaje Auditivo Visual de Rey; CN, cognitivamente normal; LMCI, deterioro cognitivo leve tardío; AD, enfermedad de Alzheimer. 36
- 3.1. Logotipo de la base de datos PPMI. 39
- 3.2. Logotipo del software Matlab. 41
- 3.3. Logotipo del software Filezilla. 41
- 3.4. Logotipo del software PuTTY. 42
- 3.5. Logotipo del software Xming. 42
- 3.6. Logotipo del software LaTeX. 43
- 3.7. Logotipo del software CeSViMa. 43
- 4.1. Procedimiento anidado de validación cruzada de 10 veces para el desarrollo y evaluación de modelos. 56
- 5.1. Valores al inicio del estudio entre sujetos PD y HC. 63
- 5.2. Correlación entre marcadores de CSF y aSyn al inicio del estudio entre PD y HC. 63
- 5.3. Valores al inicio del estudio entre sujetos PD-A+ y PD-A-. 65
- 5.4. Valores al inicio del estudio entre sujetos HC-A+ y HC-A-. 65
- 5.5. Histograma de la población PD atendiendo al tiempo de conversión y censura. En la de la izquierda, conversión a MCI o demencia. En la central, separación de conversión entre MCI y demencia. En la de la derecha, sujetos no convertidos. 68

5.6. Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores al inicio del estudio. Las líneas muestran los puntos de umbral en las medidas. 71

5.7. Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores en el análisis longitudinal. Las líneas muestran los puntos de umbral en las medidas. 71

5.8. Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color rojo para trayectoria short-term de sujetos PD-MCI+, verde para trayectoria short-term de sujetos PD-MCI- y negro para trayectoria long-term. 77

5.9. Matriz de correlación de la pendiente según el criterio SITE. Se evalúa el vector [LNS SDMT HVLTRDLY SFT MEANCAU ABETA Age] . 78

5.10. Matriz de correlación del intercepto según el criterio SITE. Se evalúa el vector [LNS SDMT HVLTRDLY SFT MEANCAU ABETA Age] . 78

5.11. Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color rojo para trayectoria short-term de sujetos PD-MCI+, verde para trayectoria short-term de sujetos PD-MCI- y negro para trayectoria long-term. 79

5.12. Matriz de correlación de la pendiente según el criterio SITE. Se evalúa el vector [SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age] . . 80

5.13. Matriz de correlación del intercepto según el criterio SITE. Se evalúa el vector [SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age] . . 80

6.1. Valores al inicio del estudio entre sujetos PD-CU, PD-MCI y PDD. . 84

6.2. Correlación entre marcadores de CSF y aSyn al inicio del estudio entre PD-CU, PD-MCI y PDD. 84

6.3. Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores al inicio del estudio. Las líneas muestran los puntos de umbral en las medidas. 85

6.4. Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores en el análisis longitudinal. Las líneas muestran los puntos de umbral en las medidas. 86

6.5. Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color verde para trayectoria short-term de sujetos PD-CU, color azul para trayectoria short-term de sujetos PD-MCI, color rojo para trayectoria short-term de sujetos PDD y negro para trayectoria long-term del modelo propuesto. 89

6.6. Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color verde para trayectoria short-term de sujetos PD-CU, color azul para trayectoria short-term de sujetos PD-MCI, color rojo para trayectoria short-term de sujetos PDD y negro para trayectoria long-term del modelo propuesto. 90

6.7. Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color verde para trayectoria short-term de sujetos PD-CU, color azul para trayectoria short-term de sujetos PD-MCI, color rojo para trayectoria short-term de sujetos PDD y negro para trayectoria long-term del modelo propuesto. 91

6.8. Trayectorias long-term de los marcadores del modelo propuesto (SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age) en una escala común, i.e en percentiles 91

A.1. Correspondencia entre hipótesis unilateral y prueba de una cola e hipótesis bilateral y prueba de dos colas 97

A.2. Indicadores de calidad. 99

A.3. Ejemplos de curvas ROC de calidades. 99

A.4. Partes de un diagrama de cajas y bigotes. 100

A.5. Relación perfecta positiva y perfecta negativa. 102

B.1. Ejemplo de gráfico interactivo de estimaciones e intervalos de comparación. 108

Índice de tablas

1.1.	Alteraciones proteicas características de cada uno de los procesos neurodegenerativos	1
1.2.	Tabla de Hoehn y Yahr.	4
2.1.	Agrupación biomarcadores AT(N). Agrupación biomarcadores AT(N) Abreviaturas: $A\beta$, β amiloide; CSF, líquido cefalorraquídeo.	13
3.1.	Distribución de los sujetos PD. PD-CU: Sujetos PD sin deterioro cognitivo. PD-MCI: Sujetos PD que desarrollan deterioro cognitivo leve. PD-Dementia: Sujetos PD que desarrollan demencia.	40
5.1.	Características demográficas y clínicas basales de los pacientes con HC y PD a los 5 años de seguimiento. Los datos son la media (desviación estándar) (mínimo, máximo) o n (%).	62
5.2.	Características demográficas y clínicas basales de los pacientes con HC-A+ y HC-A- y PD-A+ y PD-A- a los 5 años de seguimiento. Los datos son la media (desviación estándar) o n (%).	64
5.3.	Características demográficas y clínicas basales de los pacientes con PD-CU, PD-MCI y PDD a los 5 años de seguimiento con el criterio MCItest. Los datos son la media (desviación estándar) (mínimo, máximo) o n (%).	67
5.4.	Resultados de conversión a MCI o demencia respecto a los seis altos factores de riesgo (MOCA<26, APOE4, A+, PACC, Female y REM) al inicio del estudio mediante clasificación de CU, MCI y demencia.	69
5.5.	Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores al inicio del estudio y en el análisis longitudinal.	70
5.6.	Resultado del análisis de modelos predictivos univariantes al inicio del estudio con el análisis Cox y la edad como covariable.	73
5.7.	Resultado del análisis de modelos predictivos multivariantes solo con medidas neuropsicológicas al inicio del estudio con el análisis Cox y la edad como covariable.	74
5.8.	Resultados de modelos predictivos con análisis de supervivencia de Cox extendido para la población con el criterio SITE al primer año. Utilización de parámetros neuropsicológicos, del DAT-SCAN, CSF y motóricos.	74

5.9. Resultados de modelos predictivos con análisis de supervivencia de Cox extendido para la población con el criterio SITE al segundo año. Utilización de parámetros neuropsicológicos, del DAT-SCAN, CSF y motóricos.	75
5.10. Resultado del análisis de modelos predictivos multivariantes al inicio del estudio con el análisis Cox y la edad como covariable.	76
5.11. Resultados de prueba GRACE para criterio SITE.	76
5.12. Resultados del modelo de combinatoria para la obtención de modelos de progresión GRACE en Magerit	79
6.1. Características demográficas y clínicas con más discrepancias entre sujetos PD-CU, PD-MCI y PDD a los 5 años de seguimiento con el criterio SITE. Los datos son la media (desviación estándar) (mínimo, máximo) o n (%). a: discrepancia entre PD-CU y PD-MCI; b: discrepancia entre PD-MCI y PDD; c: discrepancia entre PD-CU y PDD	82
6.2. Para cada marcador seleccionado, una comparación entre los valores medios por grupo clínico (PD-CU, PD-MCI, PDD) al inicio, con respecto a los valores estimados por las trayectorias long-term al inicio de la demencia y el cutoff que se utiliza en la práctica clínica para determinar la progresión a la demencia.	89

Capítulo 1

Introducción

1.1. Neurodegeneración.

Las neuronas no se reemplazan, envejecen con la persona, asociándose a la neurodegeneración mediante la atrofia¹ compensada por la gliosis. Las neuronas no se reproducen a partir de sí mismas, ni es igual en todas las zonas del cerebro, con mayor eliminación en el hipocampo y los cíngulados [22].

La pérdida de dendritas² y la sinapsis³ afecta a la neurodegeneración. Se establece que los procesos básicos son multifactoriales, incluso ambientales y genéticos, también endógenos⁴.

La neurodegeneración se refiere a defectos en la neurona. Se puede definir como: “Producto de la acumulación de agregados de proteínas deficientemente plegadas que dan lugar a un daño oxidativo e inflamatorio que puede conducir a la muerte neuronal”.

En la tabla 1.1 se puede observar el tipo de proteína agregada, la forma de expresión y la localización en la célula.

Tabla 1.1: Alteraciones proteicas características de cada uno de los procesos neurodegenerativos

Enfermedad	Agregado proteico	Forma	Localización
Alzheimer	$A\beta$ -amiloide Tau	Placa senil Ovinilo neurofibrilar	Extracelular Citoplasma
Parkinson	α -sinucleína	C. Lewy	Citoplasma
Lewy	α -sinucleína	C. Lewy	Citoplasma
Huntington	Poliglutamina	Inclusiones	Núcleo

¹Pérdida de neuronas

²Ramificaciones de las neuronas

³Comunicación entre neuronas

⁴Anomalías de proteínas

1.2. La enfermedad de Parkinson.

La enfermedad de Parkinson (PD⁵) es un trastorno neurodegenerativo crónico de curso progresivo y evolución prolongada caracterizado por la pérdida de neuronas dopaminérgicas. Afecta entre el 4 y 5 % de los mayores de 85 años. No se conoce la causa, aunque se identifican como posibles el daño oxidativo, toxinas ambientales, la genética y el envejecimiento acelerado.

1.2.1. Síntomas.

Los síntomas son desapercibidos durante los primeros años de la enfermedad. En el comienzo, un síntoma es el temblor en reposo y asimétrico en la mano. Además, otros, como la bradicinesia o lentitud de movimientos, fatigabilidad, pérdida de destreza manual o dolor de hombros y cuello.

Se pueden describir las manifestaciones clínicas:

■ Motoras

- **Bradicinesia.** Lentitud y pérdida de movimientos espontáneos y automáticos, difícil de controlar. Es sencillo de detectar al observar la agilidad en el movimiento del paciente con movimientos diarios y simples (coger los cubiertos, atarse los cordones, etc.).
- **Temblor asimétrico en reposo.** Es el síntoma más característico de la enfermedad. Denominado “rodar de la píldora”, caracterizado por el movimiento hacia adelante y atrás del pulgar y el índice de forma rítmica. El temblor también se puede dar en otras zonas del organismo, como pueden ser manos, pies, piernas o mandíbula. Este síntoma desaparece durante el periodo de sueño o durante la realización de movimientos voluntarios.
- **Rigidez.** Resistencia muscular al movimiento pasivo en toda la amplitud del movimiento articular, sintiendo dolor y falta de flexibilidad junto con la debilidad.
- **Inestabilidad postural.** Dificultad de mantener el equilibrio, relacionada con la disminución de los reflejos posturales. Se desarrolla de forma gradual y tardía.
- **Otros síntomas motores:** disfagia⁶, síndrome de las piernas inquietas⁷ y alteraciones del lenguaje (difícilmente alzan la voz o susurrar) y/o disminución del tamaño de la letra.

■ No motoras

- **Trastornos neuropsiquiátricos.** En un tiempo más avanzado de la PD, los pacientes sufren cambios de humor espontáneos por la depresión (40-50 % de los pacientes), ansiedad (20-35 %) o apatía⁸. A esto, se le suma el efecto en las capacidades cognitivas (memoria, capacidad de atención, autocontrol emocional, etc.).

⁵ *Parkinson Disease*

⁶ Dificultad para deglutir, tragar

⁷ Deseo de mover las piernas, predominante en la noche

⁸ Falta de interés y motivación y disminución del pensamiento y respuesta efectiva.

- **Trastornos de sueño.** Se relacionan la PD con las alteraciones del sueño en la fase REM (33-46 % de los pacientes). Estos síntomas se manifiestan en forma de cuadros de insomnio o sueño fragmentado. Al final de la enfermedad, se suelen sumar alucinaciones y delirios debidos al declive cognitivo.
- **Trastornos sensoriales.** Se manifiesta la enfermedad mediante la *anosmia*⁹. Además, se pueden producir cambios en la visualización espacial, diferenciación de texturas pérdida del gusto, entre otros síntomas.
- **Otros síntomas no motores.** La PD tiene relación con problemas en vejiga o estreñimiento debidos a problemas del sistema nervioso autónomo. Además, pueden sufrir sialorrea¹⁰ y saborrea¹¹.

1.2.2. Factores de riesgo.

La causa de la aparición de la enfermedad de Parkinson es desconocida. El origen de esta enfermedad se debe a factores tanto ambientales como genéticos que se estudian para su conocimiento.

Estos múltiples factores de riesgo se clasifican en dos tipos (figura 1.1):

1. **Factores de riesgo no modificables:** Factores que son inherentes al sujeto en sí.
 - **Edad.** Este factor se debe a la pérdida de dopamina de manera fisiológica con el aumento de la edad en el circuito nigroestriado. Según los estudios se establece un 13 % de declive al año a partir de los 45 años de edad. Se establece la enfermedad de Parkinson desarrollada cuando hay una pérdida del 80 % de dopamina.
 - **Sexo.** Presenta la enfermedad una mayor frecuencia en hombres que en mujeres.
 - **Factores genéticos.** Aproximadamente, el 20 % de los pacientes diagnosticados de la enfermedad, tienen antecedentes familiares de la misma. Así, un familiar de primer grado adquiere un riesgo de padecer la enfermedad multiplicado por diez.
2. **Factores de riesgo modificables:** Factores que se pueden cambiar o corregir.
 - **Estilo de vida.** Todo lo relacionado con el estilo de vida, como puede ser la dieta, en consumo de alcohol y/o tabaco y el ejercicio físico son riesgo en la aparición o frenado del PD por su impacto en el cerebro.
 - **Traumatismo cerebral.** Los pacientes control padecen una menor ocurrencia de traumatismo craneal entre sus antecedentes que los sujetos PD.
 - **Factores ambientales.** Pesticidas, metales, campos electromagnéticos o la exposición a toxinas entre otros diversos factores ambientales aumenta el riesgo de padecer PD.

⁹Pérdida de olfato.

¹⁰Excesiva producción de saliva

¹¹Alteraciones en la piel. Seborrea en frente, lados de la nariz y cuero cabelludo.

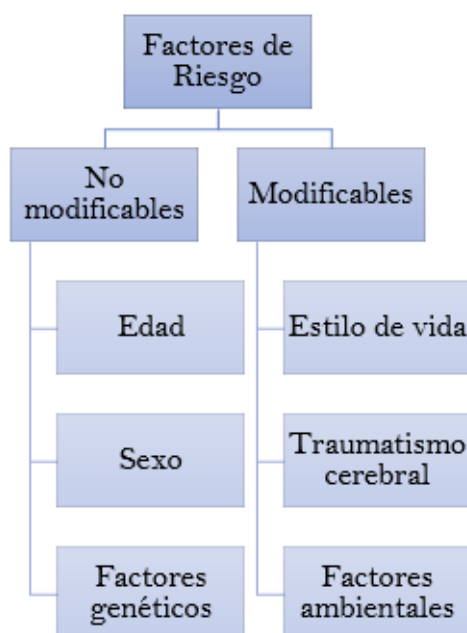


Figura 1.1: Factores de riesgo modificables y no modificables asociados al desarrollo de la enfermedad de Parkinson.

1.2.3. Diagnóstico.

El diagnóstico de la enfermedad se realiza mediante la presencia de síntomas cardinales, ausencia de signos atípicos y la respuesta al tratamiento. Es difícil realizarlo, sobre todo en las etapas iniciales. Algunas pruebas pueden ayudar, como la SPECT¹² o la PET¹³. Al igual que en el Alzheimer, se establecen escalas para definir el estadio de los pacientes. La más usada es la de Hoehn-Yahr (tabla 1.2).

Estadios evolutivos	Características
Estadio 0	Sin signos de enfermedad
Estadio 1	Síntomas y signos unilaterales
Estadio 1.5	Síntomas y signos unilaterales más afectación axial
Estadio 2	Afectación bilateral y axial
Estadio 2.5	Afectación bilateral y axial con leve recuperación en test del empujón
Estadio 3	Afectación bilateral y axial moderada con inestabilidad postural
Estadio 4	Incapacidad importante. Capaz de sostenerse y andar sin ayuda
Estadio 5	Confinado en cama o sillón. Dependencia total

Tabla 1.2: Tabla de Hoehn y Yahr.

¹²Técnica de tomografía computarizada por emisión de fotón único.

¹³Tomografía por emisión de positrones.

1.2.4. Tratamiento.

Se establecen diferentes modos de tratamiento para la Enfermedad de Parkinson.

Estrategias farmacológicas

No hay cura, pero sí medicamento que proporciona alivio de los síntomas. La levodopa es el tratamiento más eficaz. Los diferentes tratamientos son:

- **Levodopa:** Actúa como precursor metabólico de la dopamina. Retrasa el inicio de los síntomas debilitantes, sobre todo la bradicinesia y la rigidez.
- **Bromocriptina y pergolida:** Efectos secundarios como paranoia, alucinaciones, confusión, discinesias, pesadillas, náuseas y vómitos.
- **Selegilina:** Retrasa la necesidad de terapia con levodopa al menos un año.

Estrategias no farmacológicas

Mejora la calidad de vida del paciente y cuidadores. Pueden ser el ejercicio físico, la fisioterapia, terapias ocupacionales, logopedia y musicoterapia, además de educación de soporte y consejo psicosocial. Tiene tanto costes directos como indirectos:

- **Costes directos:** asistencia sanitaria y tratamientos, así como el cuidado del paciente.
- **Costes indirectos:** disminución de productividad laboral, jubilación anticipada o disminución de los cuidadores.

Estrategias quirúrgicas

Se opta por esta estrategia cuando el tratamiento farmacológico no es eficaz. Se procede a estimulación cerebral profunda en núcleos subcorticales o cirugía lesiva para solventar la enfermedad.

1.2.5. Impacto socioeconómico.

La PD es la segunda enfermedad neurodegenerativa más frecuente con una cantidad de 7 millones de personas en el mundo afectadas. El aumento de personas que padecen la enfermedad es debido al aumento de esperanza de vida. Se estima por la SEN¹⁴ que el número de afectados se triplicará en el 2050 [21]. En España, 300.000 personas aproximadamente padecen la enfermedad. De esas 300.000, el 15 % son menores de 50 años, afectando la calidad de vida.

El paciente pierde autonomía que deriva en la falta de autoestima a nivel de dependencia e impacto social. Con el progreso de la enfermedad, las limitaciones crecen con ella. Se pierde control en los movimientos, causa de los síntomas motores. Esta falta de control deriva en miedo a salir a la calle y a las caídas, aumentando la inseguridad, depresión y ansiedad. Estas tres consecuencias afectan la vida social

¹⁴Sociedad Española de Neurología

del paciente, alejándose de su vínculo más cercano.

Si hablamos de cuidados, se estima una duración de 18 horas diarias [62] las que el cuidador dedica al paciente. Esta cantidad de horas afectan al cuidador, reduciendo su tiempo propio de ocio u otras actividades personales. Esta situación deriva en el denominado "Síndrome del cuidador quemado"¹⁵. Por lo tanto, se ve que la PD afecta tanto al enfermo como al cuidador.

Se determina un coste de 17.000€¹⁶ anuales por paciente para el país. En el 2018, es una de las 15 primeras causas de muerte en España [13].

1.3. Motivación del proyecto.

Como se ha observado, hay diferentes síntomas que se manifiestan en la PD. El deterioro cognitivo es uno de los que más afecta la vida de los pacientes por su implicación en las actividades cotidianas. El deterioro cognitivo leve (MCI¹⁷) afecta al 25 % de los pacientes diagnosticados al inicio del estudio y supone un riesgo de paso a demencia [61]. La demencia afecta a un 90 % de los pacientes más avanzados.

1.4. Naturaleza del estudio.

Atendiendo la importancia del estado cognitivo del paciente PD, se hablará de tres grupos:

- PD-CU: Sujetos con Parkinson sin deterioro cognitivo.
- PD-MCI: Sujetos con Parkinson y deterioro cognitivo leve.
- PDD: Sujetos con Parkinson y demencia.

Este estudio tiene como objetivo un análisis de los pacientes en el que se evalúa el estado de los sujetos al inicio del estudio en sus diversas caracterizaciones posibles. A este estudio, se le denomina *estudio transversal*. Se estudia a todos los sujetos en un determinado estado temporal concreto. Con ello, se deben encontrar las diferencias significativas que separan a los grupos de estudio.

Por otro lado, se realiza un *estudio longitudinal* en el que se analizan los datos de los sujetos en diferentes visitas a lo largo del tiempo. Estos datos son imágenes de resonancia magnética (*MRI*¹⁸), test neuropsicológicos, test motóricos, imágenes DAT¹⁹, y datos obtenidos del CSF²⁰.

El estudio longitudinal realiza un análisis entre sujetos de la población e intra sujetos en distintos instantes de tiempo a diferencia del estudio transversal, solo con

¹⁵Sensaciones de agotamiento e impotencia causadas por la fatiga mental y el estrés.

¹⁶Estimación por la Fundación del Cerebro.

¹⁷*Mild cognitive impairment*

¹⁸Magnetic resonance imaging.

¹⁹Transporte de dopamina.

²⁰Líquido cefalorraquídeo.

análisis entre sujetos.

Se toman los datos de PPMI²¹. En estas visitas, no hay un balanceo entre ellas y/o falta información por causas ajenas a este estudio.

Se establecen modelos predictivos para combinar tanto conocimientos clínicos, biomarcadores, pruebas clínicas y datos de neuroimagen que indiquen la evolución cognitiva de un paciente y la enfermedad. Con ello, se consigue un diagnóstico preclínico a través de algoritmos de clasificación. Todo esto se realiza a través de técnicas de *Machine Learning*, que reside su fundamento teórico en la inteligencia artificial.

Por último, se busca establecer la relación entre el declive cognitivo que se produce en la PD con los constructos biológicos de la Enfermedad de Alzheimer (AD²²) mediante el análisis de las características anteriormente mencionadas.

1.5. Objetivos.

Los objetivos que se buscan alcanzar en este estudio son:

1. Estudio de forma transversal de los biomarcadores entre los tres grupos clínicos (PD-CU, PD-MCI y PDD)
 - Determinación de los biomarcadores y variables clínicas que más discriminen en el progreso del declive cognitivo de la PD.
 - Conexión con los constructos biológicos de la AD.
2. Estudio longitudinal de los biomarcadores mediante modelos predictivos.
3. Diagnóstico de la evolución cognitiva de la enfermedad en sujetos con PD.
4. Modelado temporal de la progresión del declive cognitivo en los sujetos con la PD.
5. Estimación de tiempos de conversión a MCI y demencia en los pacientes PD.

1.6. Estructura del documento

A continuación y para facilitar la lectura del documento, se detalla el contenido de cada capítulo.

- En el **capítulo 1** se realiza una introducción al proyecto tanto en el ámbito de la neurodegeneración como del Parkinson en particular. También se establecen los objetivos a alcanzar, la naturaleza e intención del estudio.
- En el **capítulo 2** se muestra un repaso al estado del arte realizado con un resumen de artículos consultados previamente y durante el estudio que han supuesto una mayor referencia en el que se indica la contribución de los mismos.

²¹Iniciativa de Marcadores de Progresión de Parkinson

²²*Alzheimer Disease*

- En el **capítulo 3** se presentan tanto los materiales y herramientas empleadas como las características de la población y el *software* empleado en el trabajo.
- En el **capítulo 4** se exponen los métodos desarrollados para alcanzar los objetivos en el trabajo.
- En el **capítulo 5** se exponen los resultados obtenidos del trabajo realizado.
- En el **capítulo 6** se estudian los resultados obtenidos y se plantan conclusiones sobre los mismos. Se manifiestan las aportaciones positivas y negativas para proponer nuevas soluciones y líneas de investigación.

Capítulo 2

Estado de la técnica.

En este capítulo se resumen los artículos que han resultado más relevantes para desarrollar el estudio y que se han empleado como referencia.

2.1. Parkinson Disease-Associated Cognitive Impairment [7].

2.1.1. Introducción

El deterioro cognitivo es hasta seis veces mayor en PD que en población sana (HC, *Healthy Control*) y es lo que más se manifiesta en la enfermedad. Este deterioro cognitivo afecta a la calidad de vida (QOL, *quality of life*) con consecuencias económicas y síntomas motores, incluso al principio, con importancia para pacientes y cuidadores. El PD-MCI se caracteriza por disminuir la capacidad cognitiva gradualmente que informa el paciente, un informante o el propio médico mediante pruebas neuropsicológicas o escalas de capacidades cognitivas globales que indican déficits cognitivos. Según el número de dominios cognitivos, podemos tener dominio único o múltiple. El PDD se produce cuando el paciente tiene deficiencia en dos o más de los cuatro dominios cognitivos, que son:

- Habilidades ejecutivas:
 - Flexibilidad mental.
 - Cambio de escenario.
 - Cambio y/o planificación eficiente del futuro.
 - Resolución de problemas.
- Atención.
- Habilidades visuoespaciales.
- Memoria.

La PDD se puede dividir en demencia leve (efecto leve a diario), moderada y grave (incapaz de llevar vida independiente). Se destaca la demencia con cuerpos de Lewy (DLB, *dementia with Lewy bodies*). Esta demencia tiene una patología de cuerpos de Lewy límbico y cortical con demencia anterior o igual a un año tras iniciar el parkinsonismo motor. DLB y PDD son muy similares, solo se distinguen en el

momento que se producen los síntomas motores y cognitivos. No existe tratamiento para prevenir o retrasar el deterioro cognitivo de la PD. Sólo existen inhibidores de colinesterasas para PDD.

2.1.2. Epidemiología

El deterioro cognitivo se produce antes o en el momento, o años más tarde de diagnosticar PD. Las personas con PD tienen un riesgo de desarrollar demencia entre 2.5 a 6 veces más que los que no lo tienen con la misma edad. No hay epidemiología clara, ya que no se incluye con frecuencia a sujetos PDD o PD-MCI, sólo en cohortes de PD establecidas. Transversalmente, entre el 24 y 31 % de sujetos PD tienen demencia. Entre 54 y 70.2 años, la prevalencia de PDD es:

- 17 % a los 5 años de diagnosticarlo.
- 46 %, a los 10 años.
- 83 % a los 20 años.

Se puede observar gran variabilidad en la llegada de la demencia. En cambio, en la población general de más de 60 años de edad, la prevalencia es de 5 al 7 %. En la última década se presta más atención al MCI, como en el caso de la AD. El 25.8 % de los sujetos con PD sin demencia tienen MCI. Aproximadamente, el 20.2 % de los pacientes tienen MCI cuando se diagnostican (71.3 ± 7.5) y aumenta al 40-50 % tras 5 años. En cambio entre los 60 y 90 años de edad de HC prevalece entre el 16 al 20 %.

El MCI se describe como etapa transitoria entre cognición normal (PD-NC, *Parkinson Disease - normal cognition*) y la demencia. El riesgo a largo plazo de demencia es mayor en pacientes con PD que revierten de MCI a PD-NC. También surgen en el campo de AD.

Se pueden establecer los siguientes predictores de mayor a menor orden de peso asociados con el desarrollo de deterioro cognitivo o demencia:

- Presencia de alucinaciones.
- Edad avanzada.
- Síntomas de gravedad general del trastorno motor.
- Presencia de deterioro del habla.
- Mayor edad al inicio del PD.
- Gravedad de la bradicinesia.
- Mayor estadio de Hoehn y Yahr (5 etapas para describir la gravedad de PD).
- Deterioro axial (postura inestable, marcha con dificultades, etc.).
- Bajo nivel educativo.

- Presencia de depresión.
- Sexo masculino.

Características prodómicas como hiposmia (pérdida de olfato), trastorno del comportamiento del sueño REM (RBD) y reducción de la unión al transportador de dopamina, pueden tener peores síntomas. El déficit cognitivo se ha establecido como nuevo marcador prodómico en la investigación del PD prodómica. La neuropatología más común en PDD es la patología de Lewy límbica y/o neocortical. Las patologías relacionadas con la edad más comunes en PD y deterioro cognitivo son las de AD, el depósito de β -amiloide extracelular y el depósito intracelular de tau, distribuidas de forma diferente a las patologías de Lewy. Un tercio de los pacientes con PDD tiene PET de β -amiloide positiva, con prevalencia de MCI a AD a lo largo del tiempo.

2.1.3. Diagnóstico, cribado y prevención

Se puede realizar a partir de evaluación de la función cognitiva global o detallada que evalúa:

- Atención.
- Memoria de trabajo.
- Función ejecutiva.
- Función del lenguaje.
- Función de la memoria.
- Función visuoespacial.

El cribado es de menos tiempo, menos recursos, más disponible y menos costoso que las evaluaciones neuropsicológicas detalladas. Presentan desventajas como la información limitada del perfil cognitivo y confiabilidad reducida. En base a propiedades clinimétricas en PD, se recomendaron tres escalas para el cribado:

- Evaluación cognitiva de Montreal (MoCA): El punto de corte óptimo de 23/24 tiene sensibilidad de 0.41 y especificidad de 0.82, con 68% de diagnóstico correcto en PD-MCI.
- *Mattis Dementia Rating Scale Second Edition* (MDRS2).
- Enfermedad de Parkinson - Escala de calificación cognitiva (PDCRS).

También se establecen dos escalas recomendadas con salvedades:

- *MiniMental Parkinson* (MMP).
- Escala de resultados en la cognición de la enfermedad de Parkinson (SCOPA-COG).

Las pruebas neuropsicológicas son pruebas estandarizadas validadas con normas poblacionales adecuadas. Están influenciadas por la edad y el nivel educativo. Se dividen las pruebas en 5 dominios:

- Atención y memoria de trabajo.
- Ejecutivo.
- Lenguaje.
- Memoria .
- Funciones visoespaciales.

Para clasificar PD-MCI del MDS como SCD se realiza una entrevista con el paciente distinguiendo MCI de demencia mediante PD-CFRS¹ o PDAQ-15².

Hay dos niveles de PD-MCI:

- Nivel I: Evaluación abreviada. Ejemplo: Detección de función cognitiva o batería de pruebas psicológicas.
- Nivel II: Pruebas neuropsicológicas integrales de cinco dominios cognitivos.

Para diferenciar, de manera tradicional, si la demencia tiene más de un año desde el diagnóstico de PD, es PDD. Si es desde el momento de diagnóstico se clasifica como DLB. Al principio de PD, hay baja concentración de α -sinucleína, relacionada con su inclusión en la formación de cuerpos de Lewy y seguida por concentraciones aumentadas debidas a la fuga de α -sinucleína que reflejan la mayor neurodegeneración.

2.2. Toward a biological definition of Alzheimer's disease [6].

2.2.1. Introducción

Existen biomarcadores utilizados en el Alzheimer como sustitutos de los cambios neuropatológicos. Por eso, se puede basar en biomarcadores de personas vivas para definir esta enfermedad. En esta investigación se trata de caracterizar a los participantes con los biomarcadores β -amiloide ($A\beta$), tau patológica y neurodegeneración. El deterioro cognitivo en el Alzheimer se produce durante un largo período de forma continua al igual que los biomarcadores anteriores a los síntomas, dejando atrás las tres entidades clínicas: MCI, Enfermedad del Alzheimer preclínica y demencia. Se recomienda usar biomarcadores de imágenes y de líquido cefalorraquídeo (CSF). En pacientes con síntomas se utilizan biomarcadores para confirmar los cambios patológicos en el deterioro cognitivo. Se establecen a los amiloides como los biomarcadores en la cúspide de la jerarquía de biomarcadores preclínicos, mientras que el resto están en igualdad de MCI y demencia. Por otro lado, se ha demostrado de que las medidas de lesión neuronal como MRI (imágenes por resonancia magnética), FDG (PET con florodesoxiglucosa) y T-tau (tau total en CSF) no son específicas de la enfermedad, sino derivaciones etiológicas.

¹Parkinson's Disease-Cognitive Functional Rating Scale

²Penn Parkinson's Daily Activities Questionnaire-15

2.2.2. Alzheimer como cambios neuropatológicos

Se especifica que la demencia es un síndrome compuesto por signos y síntomas de enfermedades entre las que se encuentra el Alzheimer.

Entre el 10% y 30% de las personas con demencia de Alzheimer no muestran cambios neuropatológicos en la autopsia de esta enfermedad. Proporciones similares tienen amiloide normal PET o CSF $A\beta$, por lo que no se encuentra con un fenotipo de demencia amnésica multidominio, pudiendo ser así fruto de más enfermedades además del Alzheimer. Otros trastornos como los del lenguaje, visuoespaciales y ejecutivos pueden deberse de ello. Los cambios neuropatológicos no se suelen reflejar en signos o síntomas, predominando en edad avanzada (entre el 30-40% de personas mayores sin deterioro cognitivo (CU) tienen cambios neuropatológicos de Alzheimer en la autopsia, proporción similar en biomarcadores amiloides anormales). Con esto se establece que los síntomas cognitivos no son la mejor forma de definir a la enfermedad.

Se utilizan biomarcadores de imágenes y CSF para investigar en el Alzheimer. Se reconocen tres grupos generales de biomarcadores basados en la naturaleza del proceso patológico que mide cada uno.

- Los biomarcadores de placas $A\beta$ (etiquetadas en “A”) son ligando de PET amiloide cortical que se une o CSF bajo $A\beta_{42}$.
- Los biomarcadores de tau fibrilar (etiquetados como “T”) son tau fosforilada en CSF elevada (P-tau) y PET tau cortical la unión de ligando.
- Los biomarcadores de neurodegeneración o lesión neuronal (etiquetados como “N”) son CSF-tau, hipometabolismo FDG PET y atrofia en la resonancia magnética.

Tabla 2.1: **Agrupación biomarcadores AT(N)**. Agrupación biomarcadores AT(N) Abreviaturas: $A\beta$, β amiloide; CSF, líquido ceforraquídeo.

Agrupación biomarcadores AT(N)	
A: Agregado $A\beta$ o estado patológico asociado	CSF $A\beta_{42}$, o $A\beta_{42}/A\beta_{40}$ relación PET amiloide
T: tau agregada (ovillos neurofibrilares) o estado patológico asociado	Tau fosforilada en CSF PET Tau
(N): Neurodegeneración o lesión neuronal	RM anatómica PET FDG Tau total en CSF

Según la NIA-AA (2011), los biomarcadores se agrupan en: neurodegeneración amiloide y relacionada con tau. La tauopatía y la neurodegeneración están en la misma categoría de biomarcadores. En personas con solo Alzheimer, se asocia estrechamente la neurodegeneración con la tau patológica pero, también se produce en condiciones distintas a esta enfermedad.

Si lo clasificamos como AT(N), separando los biomarcadores específicos para la tau patológica de las medidas inespecíficas de neurodegeneración, solucionamos el problema anterior. Esta clasificación se diseñó con CSF y un biomarcador en cada grupo (tabla 2.1), pudiéndose caracterizar estos biomarcadores solo de imágenes o CSF. Otros prefieren combinación de imágenes y biomarcadores de CSF.

2.2.3. Definición de la Enfermedad del Alzheimer

Solo los biomarcadores distintivos del Alzheimer, $A\beta$ y tau patológico, deben considerarse como definiciones de biomarcadores de la enfermedad. Los pacientes con CU y biomarcadores amiloides anormales progresan más rápido de atrofia, hipometabolismo y deterioro cognitivo que los que no evidencian biomarcadores $A\beta$. La proporción de individuos normales clínicamente con PET amiloide positivo por edad se asemeja crecientemente a la prevalencia por edad diagnosticados como demencia 15-20 años después. Se define:

- “Alzheimer continuo”: evidencia de solo biomarcadores $A\beta$ con un biomarcador tau patológico normal.
- “Enfermedad de Alzheimer”: evidencia de biomarcadores $A\beta$ y tau patológico está presentes.

2.2.4. Puesta en escena

Dos tipos de información del participante se organizan de forma independiente:

1. Clasificación de la gravedad de la enfermedad mediante biomarcadores.
2. Clasificación de la gravedad del deterioro cognitivo.

Las medidas para definir la enfermedad deben ser específicas, mientras que las que determinan la gravedad, no.

- $A\beta$ determina si el individuo se encuentra o no en el continuo de la enfermedad el Alzheimer.
- Los biomarcadores patológicos de tau determinan si alguien con Alzheimer continúa en la enfermedad.
- Los biomarcadores de lesión neurodegenerativa/neuronal y síntomas cognitivos, de los cuales ninguno es específico para la enfermedad, pero sí para la gravedad.

2.2.5. Perfiles y categoría de biomarcadores

Si añadimos un punto de corte normal/anormal para cada grupo de biomarcadores AT(N), da como resultado ocho AT (N) diferentes, “perfiles de biomarcadores”, que son: A+T-(N)-, A+T+(N)+, etc.

Basado en el cambio patológico de Alzheimer y enfermedad e Alzheimer, es sistema ATN asigna a cada individuo a uno de tres categoría de biomarcadores:

1. Individuos con biomarcadores AD normales.
2. Aquellos en el continuo de Alzheimer, subdivididos en:
 - Cambio patológico de Alzheimer
 - Enfermedad de Alzheimer
3. Con biomarcador amiloide normal pero con T o (N) anormal, o ambos. Etiquetada como fisiopatología sospechosa no relacionada con el Alzheimer (SNAP).

La puntuación AT(N) es independiente de los síntomas clínicos.

- A-T-(N)-: Estado sin evidencia de cambio patológico detectable por biomarcadores AT(N),
- A+T+(N)+: Estado patológico avanzado.

A mayor número de grupos de biomarcadores anormales, más avanzado será el estadio patológico.

2.2.6. Características y limitaciones de los biomarcadores

- Biomarcadores CSF: medidas de las concentraciones de proteínas en el CSF del saco lumbar reflejando las tasas de producción y aclaramiento.
- Medidas de imagen: Representan la magnitud de la carga o daño neuropatológico acumulados a lo largo del tiempo.

Bajo CSF $A\beta_{42}$ se considera mejor biomarcador de un estado patológico con la formación de placa amiloide y no con la carga de placa amiloide como es la PET amiloide. De manera similar, la P-tau en CSF se considera como biomarcador de un estado patológico asociado a la formación de PHF tau y no de depósitos patológicos de tau como es tau PET. Puede haber discordancias entre imágenes y biomarcadores de CSF por los puntos de corte, y que se pueden rectificar ajustándolos. La relación entre CSF $A\beta_{42}$ y PET amiloide tiene forma de "L" por un desplazamiento temporal entre las dos medidas. La unión del ligando de tau PET se correlaciona linealmente con P-tau en CSF, pero de forma imperfecta debido a la estabilización más adelante de P-tau en la enfermedad, mientras tau PET sigue aumentando.

2.2.7. CSF T-tau y P-tau

El aumento de T-tau sin cambios en P-tau en la lesión cerebral traumática y el accidente cerebrovascular se correlaciona con la gravedad de daño neuronal. El único trastorno que muestra aumento de la P-tau en el CSF es en la Enfermedad de Alzheimer, siendo normal en otros trastornos neurodegenerativos. También, P-tau en CSF se correlaciona con la grave acumulación de PHF tau después de la muerte. Todos estos datos indican que CSF T-tau nos indica la intensidad de daño neuronal en un punto, mientras que la P-tau en CSF indica un estado patológico anormal

asociado con la formación de PHF tau.

Estos biomarcadores no son específicos de la neurodegeneración en el Alzheimer. Se desconoce la proporción neurodegeneración/lesión debida a la enfermedad frente a otras posibilidades. Se diferencia de A y T, en que no se asignan los biomarcadores a los hallazgos neuropatológicos, por eso se establece entre paréntesis: (N).

2.3. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study [2].

2.3.1. Introducción

En el 75 % de los enfermos de Parkinson durante más de 10 años ocurre la demencia, el deterioro de la cognición que los incapacita. Nos referimos al deterioro cognitivo leve para aquel deterioro cognitivo que no cumple requisitos de demencia en el Parkinson. La mayoría de los pacientes con deterioro cognitivo leve evolucionan a demencia.

Existen varios mecanismos para detectar el deterioro cognitivo, detectando la coexistencia del Alzheimer (placas amiloides $A\beta$ y ovillos neurofibrilares tau) y la patología de la enfermedad de Parkinson (cuerpos corticales de Lewy). El déficit de dopamina refleja la mejora de síntomas cognitivos al principio debido a la levodopa. La neuroimagen asocia la densidad del transportador de dopamina caudado y putamen con la disfunción cognitiva. La edad avanzada, la menor educación, el sexo, puntuación cognitiva, graves síntomas motores, hiposmia y trastorno del comportamiento del sueño (RBD) con movimiento oculares rápidos (REM) predicen el deterioro cognitivo en pacientes con Parkinson.

2.3.2. Métodos

Se investigan predictores clínicos y biomarcadores del deterioro cognitivo en la primera etapa utilizando datos de la *Parkinson's Progression Marker Initiative* (PPMI). Los criterios de inclusión son:

- Tener más de 30 años.
- Temblor asimétrico en reposo o bradicinesia asimétrica o dos de los tres signos de bradicinesia.
- Temblor en reposo y rigidez.
- Diagnóstico reciente de enfermedad de Parkinson.
- Estar en la etapa 1 o 2 de Hoehn y Yahr.
- No recibir medicamentos para el Parkinson 60 días después de la primera visita.
- No requiere medicación dentro de al menos 6 meses desde el inicio.

2.3. CLINICAL VARIABLES AND BIOMARKERS IN PREDICTION OF COGNITIVE IMPAIRMENT IN PATIENTS

- Déficit de DAT en las imágenes.
- No tener medicamentos con interferencia en DAT o CSF.
- No haber dispuesto de medicamentos o dispositivos en investigación 6 días después de la primera visita.

Los criterios, por otro lado, de exclusión son:

- Trastornos neurológicos clínicamente significativos.
- Pariente de primer grado con enfermedad de Parkinson.
- Puntuación de MoCA (Evaluación Cognitiva de Montreal) menor o igual a 26.
- Ingesta de medicamentos interfiriendo imágenes DAT o impedir la correlación de CSF.
- Haber dispuesto de medicamentos o dispositivos en investigación 6 días después de la primera visita.

2.3.3. Resultados y Discusión.

Evaluación del deterioro cognitivo por el cambio de MoCA, escala de capacidades cognitivas globales validadas en Parkinson desde el inicio y durante 2 años, evaluación de MoCA en 2 años y categorización como deterioro cognitivo a los 2 años de seguimiento.

Las pruebas cognitivas para la función cognitivas fueron:

- HVLT-R: Prueba de Aprendizaje Verbal de Hopkins revisada para la memoria.
- Versión *Benton Judgment of Line Orientation* de 15 elementos (mitad dividida) para la función visuoespacial.
- Prueba de Modalidades Símbolo-Dígito para la velocidad.
- Secuencia de número y letras para la función ejecutiva y memoria de trabajo.
- Prueba de fluidez semántica.

Con ello se clasificaron en (criterio de protocolo PPMI):

- Cognición normal.
- Deterioro cognitivo leve: puntuaciones en dos o más de la memoria total JvLR, discriminación de reconocimiento HVLT, discriminación de reconocimiento HVLT, *Benton Judgment of Line Orientation*, secuenciación de letras y números, prueba de fluidez semántica o de símbolos y dígitos con más de 1.5 desviaciones por debajo de lo normal debido al deterioro cognitivo.
- Demencia: además de lo anterior, deterioro funcional en la vida diaria.

Otras variables incluidas fueron:

- Edad y sexo.
- Años de educación.
- Duración de la enfermedad.
- Sentido del olfato mediante la UPSIT (Prueba de identificación del olfato de la Universidad de Pensilvania).
- RBD mediante el RBDSQ (Cuestionario de detección de RBD).
- Depresión medida por la GDS (Escala de depresión geriátrica) de 15 items.
- Gravedad de la enfermedad mediante la puntuación motora de la MDS-UPDRS (Escala Unificada de Calificación de la Enfermedad de Parkinson de la Sociedad de Trastornos del Movimiento).
- Subtipo de temblor dominante.
- Subtipo de inestabilidad postural.
- Dificultad para caminar.
- Subtipo motor indeterminado.

Para los estudios de biomarcadaores se incluyen:

- Datos de imágenes DAT para captación media de caudado y putaminal en relación con la captación en el área occipital. Asimetría de captación de caudado y putaminal (lado con mayor dividido por lado con menor captación)
- APOE Estado $\epsilon 4$.

a. Análisis estadístico

La puntuación MoCA se calculó como la diferencia en la línea de base de MoCA y las puntuaciones de 2 años, buscando datos faltantes. Comparamos grupos con x^2 test, t test para variables normalmente distribuidas y Mann-Whitney test para datos no paramétricos. Se comprueba los residuos, cumpliendo con los supuestos de regresión lineal, sin sesgarse negativamente en el modelo final, y verificándose gráficamente y con pruebas de normalidad, además de la homocedasticidad y la independencia.

Se realizan análisis lineales univariados y multivariados con puntuaciones de MoCA entre el inicio y 2 años, teniendo la puntuación de MoCA a los dos años como variables dependientes en la regresión lineal hacia atrás. Las variables con $p < 0.2$ en modelos univariados se llevan a eliminación hacia atrás con p -remoción = 0.1.

Si hay variables con una correlación alta, la variable con p más bajo se estableció como independiente. No se incluye la puntuación MoCA al inicio en los análisis de regresión lineal multivariante para predecir su cambio a 2 años y así no incluirlos en

ambos lados de la ecuación de regresión.

Para el modelo con variables clínicas y biomarcadores, solo se incluyen variables independientes con $p < 0.005$. Se aplica un remuestreo bootstrap con 1000 repeticiones al modelo de riesgo final, confirmando la precisión de la predicción mediante la validación cruzada de diez veces y división de cohortes.

Para predecir el riesgo de pacientes, se construye un modelo de riesgo para calcular el riesgo predicho:

$$\text{riesgo de deterioro cognitivo del paciente a los 2 años} = 1 + \exp[\text{puntuación de riesgo del paciente}]$$

Donde:

- Puntuación de riesgo del paciente = intersección + (variable b1 x variable 1) + (variable b2 x variable 2) + (variable b3 x variable 3) + (variable b4 x variable 4) + (variable b5 x variable 5)
- Variable b1, variable b2, variable b3, variable b4 y variable b5 son los coeficientes de regresión que describen cómo los valores de las variables de un paciente afectan al riesgo.

b. Resultados

Tres individuos no tenían datos MoCA basales y fueron excluidos. Las características basales se muestran en la tabla 2.1. No hay diferencias significativas en las variables evaluadas entre individuos con y sin datos de seguimiento cognitivo de 2 años, excepto la asimetría caudada media (mayor en los que no tienen evaluación cognitiva a los 2 años, $p=0.02$). Imputamos los datos faltantes con medias, no alterando los resultados.

En análisis multivariados de los predictores de cambio en la función cognitiva, el cambio en las puntuaciones de MoCA desde el inicio hasta los dos años se asocia a la edad, puntuación UPSIT, estado APOE y CSF $A\beta_{42}$ a relación t-tau. En cambio, en pacientes sanos la puntuación MoCA se asocia con la edad, sexo y CSF $A\beta_{42}$.

El 16 % de los participantes fueron clasificados con deterioro cognitivo leve, un 1 % fue clasificado como demencia y un 83 % como cognitivamente saludables. Un 1 % de los controles sanos se clasificaron como deterioro cognitivo leve y ninguno tenía demencia a los dos años. A los dos años, 89 % de los pacientes sin deterioro cognitivo y 92 % con deterioro se trató con medicación antiparkinsoniana ($p=0.47$). El 8 % de los pacientes con deterioro recibieron medicación para el deterioro cognitivo. Esos pacientes con Parkinson que se clasificaron con deterioro cognitivo, eran mayores, con altas puntuaciones de RBDSQ más altas en el inicio del estudio, puntuaciones UPSIT más bajas, $A\beta$ en CSF más bajo y $A\beta_{42}$ a t-tau, y menor captación media del caudado y asimetría caudado y putaminal que aquellos que no tenían deterioro.

Con análisis de regresión logística del deterioro como variable dependiente, teniendo como independientes la edad y otras variables clínicas ($p < 0.05$), la precisión fue mayor que para solo la edad. Los resultados de la validación cruzada 10 veces confirman el modelo final sin diferencias en el rendimiento.

Se identifican predictores clínicos y biomarcadores de deterioro cognitivo en los primeros dos años después de diagnosticarse Parkinson, predictor de la demencia en esta enfermedad, asociado a variables clínicas independiente de las puntuaciones MoCA relativas o absolutas o una clasificación del deterioro cognitivo.

2.4. Brain Amyloid Contribution to Cognitive Dysfunction in Early-Stage Parkinson's Disease: The PPMI Dataset [16].

2.4.1. Introducción.

La disfunción cognitiva destaca en los síntomas motores de la Enfermedad del Parkinson (Parkinson's disease, PD), progresando de forma significativa hasta la demencia y afectando negativamente a la vida de las personas que la padecen.

El deterioro del rendimiento cognitivo de la PD se debe a:

- Cuerpos de Lewy corticales y límbicos.
- Degeneración de neuronas colinérgicas de prosencéfalo basal.
- Pérdida desigual de dopamina en neuronas nigroestriatales.
- Placas β -amiloide ($A\beta$).
- Ovillos neurofibrilares tau.

La medición de la captación de amiloide cerebral por PET y los niveles de amiloide en el líquido cefalorraquídeo (CSF) se establecen como posibles biomarcadores de demencia en PD. Al morir, la mitad de los pacientes con PD o demencia con cuerpos de Lewy muestran patología amiloide que respaldan la enfermedad de Alzheimer. Amiloide y tau de forma conjunta y en alta concentración disminuye la supervivencia y aumenta la demencia.

El papel del amiloide en la etapa temprana de la PD es escasa, basándose en CSF. Hay estudios de PET con PIB (Compuesto B de Pittsburgh) que muestran depósito amiloide en pacientes con PD, sin evaluar la etapa temprana. Este estudio pretende investigar asociación entre unión de fluorbetaben de $A\beta$, niveles de amiloide en CSF y disfunciones cognitivas en la PD temprana de la base de datos de PPMI, suponiendo que el aumento de la captación de amiloide cerebral modula manifestaciones cognitivas tempranas.

2.4.2. Métodos

. Partiendo de PPMI, se investigan datos clínicos, cognitivos y de neuroimagen a marzo de 2018 de 87 pacientes con florbetaben PET. Se excluyen 14 controles sanos, 16 prodrómicos de PD y 9 sin cohorte genético, dando un total de 48 pacientes con PD sintomática. Se incluyen características neuropsicológicas y clínicas y resultados de CSF.

2.4.3. Características neuropsicológicas y clínicas.

Las variables demográficas y clínicas incluyeron:

- Edad.
- Años de educación.
- Sexo.
- Duración de la enfermedad.
- Dosis diaria equivalente de levodopa.
- Dosis diaria equivalente de agonista dopaminérgico.
- Gravedad de la enfermedad medida mediante MDS-UPDRS III (Encendido/Apagado)
- Funcionamiento diario usando las ALD.

Así, se clasifican los pacientes PD en:

- Manifestaciones de temblor dominante (TD).
- Inestabilidad postural-alteración de la marcha (PIGD).
- Fenotipo de PD intermedio.

La cognición global se evalúa con MoCA en el examen de neuroimagen. En cambio, las funciones cognitivas específicas mediante:

- Dominio visuoespacial por JLO (*Benton de Orientación de Línea*).
- Atención, exploración visual y velocidad motor por el SDMT (Prueba de modalidades de símbolo-dígito).
- Recuerdo inmediato y retardo de la memoria con HVLT-R (Prueba de Aprendizaje Verbal de Hopkins-Revisada).
- Atención y memoria de trabajo con la LNS (secuenciación de letras y números).
- Habilidades del lenguaje con la prueba de fluidez semántica.

Se establece una submuestra de los participantes ($n=44$) en el que nos encontramos: 10 participantes $PDA\beta+$ y 34 $PDA\beta-$. Esta, se somete a punción lumbar inicialmente (2.5 ± 0.8 años antes de la PET amiloide) para obtener muestras de CSF. Para reducir variabilidad, el inmunoensayo Roche Elecsys (Roche Diagnostic) analiza $A\beta_{1-42}$, concentraciones de t-tau (tau total) y p-tau (tau fosforilada). Por otra parte, el ensayo de BioLegend, inmunoabsorbente ligado a enzimas, mide las concentraciones de α -sinucleína.

Las imágenes se adquirieron en centros PPMI con protocolo de imágenes estándar. Las imágenes PET de florbetaben se importaron al software de cuantificación de imágenes biomédicas PMOD para procesarlas y analizarlas con un control de calidad técnico realizado en un laboratorio central de imágenes.

Se evaluó en movimiento de los marcos dinámicos de PET y la corrección necesaria. Estos archivos se promedian creando un solo volumen de PET. Ese volumen se normaliza al espacio estándar del MNI para coincidir con todas las exploraciones en la alineación anatómica. El PET normalizado se convierte en valores de captación estándar (SUV). Los volúmenes de interés (VOI) de la plantilla de etiquetado anatómico automatizado modificado de MNI, con regiones corticales y subcorticales se aplicaron al volumen de PET de SUV y se ajustan a la atrofia del sujeto. La ubicación de VOI se guarda para cada sujeto.

Se extraen medidas semicuantitativas (SUB promedio por vóxel) para calcular las relaciones SUV regionales (SUVR), utilizando corteza cerebrosa como región de referencia. Se establece un SUVR compuesto a cada sujeto con el cálculo de los SUVR medios de las regiones de interés.

- Positivos ($PDA\beta+$): SUVR compuestos > 1.43 .
- Negativos ($PDA\beta-$): SUVR compuestos < 1.43 .

La distribución de la muestra no es normal. Se realizan estadísticas descriptivas y no paramétricas, analizando datos demográficos, clínicos y de imagen. Las variables categóricas se comparan con Fisher. Luego, correlaciones de rango de Spearman para la asociación cognición/deposiciones amiloide. El análisis estadístico se realizó con IBM SPSS Statistics v20.0 con una significancia del 5%. También se utilizó la FDR para corregir comparaciones múltiples.

2.4.4. Resultados y discusión.

Entre los 48 PD, 10 (21%) tiene florbetaben PET SUVR > 1.43 y se clasificaron como $PDA\beta+$, mientras que los otros 38 fueron $PDA\beta-$. Los datos clínicos se presentan en la tabla 1 del artículo (figura 2.1). Se destaca la puntuación MoCA y SDMT, que es más baja en pacientes $PDA\beta+$ que en $PDA\beta-$. También se observan diferencias en el CSF p-tau/ $A\beta_{1-42}$ ($p<0.002$) y T-tau/ $A\beta_{1-42}$ ($p<0.001$).

La imagen correspondiente a la tabla 2 (Figura 2.2), muestra SUVR regionales de florbetaben en $PDA\beta+$ versus $PDA\beta-$. Se manifiesta aumento muy significativo

Table 1
Demographic, clinical and neuropsychological characteristics according to amyloid status

	PDAB ⁻ (n = 38)			PDAB ⁺ (n = 10)			p
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	
Age, y	63.97 (8.89)	49.00	84.00	70.40 (8.54)	55.00	81.00	0.062
Education, y	16.24 (2.71)	12.00	26.00	17 (2.11)	14.00	20.00	0.248
Sex (m/f)	26/12			8/2			0.701
Disease duration, m	3.76 (1.41)	1.00	8.00	3.98 (1.63)	1.59	7.51	0.750
LEDD	424.24 (236.23)	50.00	1100.00	282.25 (148.77)	105.00	562.50	0.090
DAED	83.53 (168.81)	0	825.00	30.50 (50.03)	0	120.00	0.556
MDS- UPDRS III (on)	23.79 (11.00)	5.00	56.00	25.10 (9.24)	17.00	48.00	0.866
MDS- UPDRSIII (off)	29.96 (10.29)	10.00	51.00	29.33 (4.46)	24.00	35.00	0.750
Motor phenotype							
TD/PIGD/IND	24/11/3			6/4/0			0.574
TD subscore	5.18 (3.59)	0	13.00	5.90 (4.61)	0	12	0.680
PIGD subscore	1.58 (0.98)	0	5.00	2.00 (1.49)	0	4.00	0.415
CSF markers, pg/mL _a							
Aβ ₁₋₄₂	1002.29 (367.65)	516.90	1700.00	837.63 (348.07)	497.20	1373.00	0.202
p-tau181	13.72 (5.11)	8.00	29.63	18.06 (10.42)	10.07	45.90	0.098
T-tau	164.62 (58.25)	80.00	299.70	208.66 (105.56)	107.00	475.20	0.218
α-Syn	1505.55 (830.64)	638.9	4954.90	1609.41 (959.65)	660.70	4105.30	0.695
T-tau/Aβ ₁₋₄₂ ratio	0.17 (0.05)	0.10	0.33	0.28 (0.21)	0.16	0.85	0.001
p-tau/Aβ ₁₋₄₂ ratio	0.01 (0.004)	0.01	0.03	0.02 (0.02)	0.01	0.08	0.002
p-tau/T-tau ratio	0.08 (0.01)	0.07	0.10	0.09 (0.01)	0.07	0.10	0.363
Cognitive tests							
MoCA	27.11 (2.25)	21.00	30.00	24.80 (2.62)	22.00	29.00	0.013
LNS	11.76 (2.67)	8.00	18.00	10.80 (3.23)	6.00	16.00	0.459
SDMT	48.60 (8.96)	29.17	66.25	41.19 (9.33)	30.00	56.67	0.030
Benton JLO	12.50 (2.47)	7.12	16.38	10.72 (2.98)	6.20	13.44	0.233
HVLT-R immediate recall	51.55 (10.05)	33.00	71.00	49.80 (8.30)	38.00	60.00	0.721
HVLT-R delayed recall	50.13 (8.62)	36.00	64.00	46.40 (13.44)	25.00	61.00	0.502
Semantic fluency	52.87 (10.24)	35.00	80.00	50.90 (10.83)	38.00	69.00	0.479
GDS	4.89 (1.24)	1.00	7.00	4.90 (1.52)	2.00	7.00	0.898
ADL	87.24 (8.20)	70.00	100.00	89.5 (7.62)	80.00	100.00	0.457

PD, Parkinson's disease; Aβ, amyloid-β; SD, standard deviation; LEDD, levodopa equivalent daily dose; MDS-UPDRS, Movement Disorder Society Unified Parkinson's Disease Rating Scale; TD, tremor dominant; PiGD, postural instability-gait disturbance; IND, intermediate; CSF, cerebrospinal fluid; T-tau, total-tau, p-tau, phosphorylated-tau; α-Syn, α-synuclein; MoCA, Montreal Cognitive Assessment; LNS, Letter Number Sequencing test; SDMT, Symbol Digit Modalities Test; JLO, Judgement of Line Orientation; HVLT-R, Hopkins Verbal Learning Test-Revised; GDS, Geriatric Depression Scale; ADL, Activity of Daily Living; Significant value is in bold type. a: CSF data was available for a subset of PDAB⁻ (n = 34) and for all PDAB⁺.

Figura 2.1: Tabla 1. Características demográficas, clínicas y neuropsicológicas de acuerdo con el estado amiloide.

en regiones como la corteza , núcleos subcorticales y protuberancia.

Table 2
Regional [¹⁸F]florbetaben SUVRs uptake by amyloid status

Region	PDAβ- (n = 38)			PDAβ+ (n = 10)			p
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	
Frontal R	1.29 (0.08)	1.15	1.49	1.67 (0.30)	1.41	2.37	<0.001
Frontal L	1.31 (0.08)	1.16	1.49	1.67 (0.28)	1.41	2.29	<0.001
Orbitofrontal R	1.24 (0.09)	1.08	1.44	1.54 (0.29)	1.25	2.13	<0.001
Orbitofrontal L	1.24 (0.08)	1.10	1.43	1.51 (0.25)	1.29	2.01	<0.001
Rectus R	1.20 (0.11)	0.94	1.41	1.56 (0.34)	1.13	2.31	<0.001
Rectus L	1.24 (0.09)	1.05	1.47	1.55 (0.29)	1.15	2.19	<0.001
Temporal R	1.23 (0.07)	1.10	1.38	1.48 (0.17)	1.33	1.89	<0.001
Temporal L	1.22 (0.08)	1.00	1.46	1.45 (0.12)	1.29	1.71	<0.001
Mesial temporal R	1.22 (0.07)	1.08	1.38	1.41 (0.13)	1.31	1.72	<0.001
Mesial temporal L	1.22 (0.08)	1.07	1.48	1.39 (0.10)	1.28	1.62	<0.001
Lateral temporal R	1.24 (0.07)	1.10	1.38	1.50 (0.19)	1.33	1.94	<0.001
Lateral temporal L	1.23 (0.08)	1.11	1.45	1.48 (0.13)	1.30	1.75	<0.001
Anterior cingulate R	1.25 (0.11)	1.04	1.55	1.67 (0.38)	1.32	2.45	<0.001
Anterior cingulate L	1.30 (0.12)	1.05	1.52	1.74 (0.33)	1.40	2.47	<0.001
Posterior cingulate R	1.30 (0.13)	1.07	1.54	1.75 (0.29)	1.32	2.38	<0.001
Posterior cingulate L	1.37 (0.12)	1.10	1.56	1.83 (0.25)	1.56	2.37	<0.001
Parietal R	1.26 (0.09)	1.09	1.44	1.63 (0.26)	1.38	2.22	<0.001
Parietal L	1.26 (0.09)	1.08	1.43	1.62 (0.24)	1.38	2.09	<0.001
Occipital R	1.30 (0.08)	1.16	1.52	1.52 (0.09)	1.41	1.67	<0.001
Occipital L	1.32 (0.08)	1.18	1.52	1.57 (0.10)	1.44	1.73	<0.001
Caudate R	1.41 (0.11)	1.11	1.68	1.70 (0.27)	1.40	2.26	<0.001
Caudate L	1.35 (0.12)	1.12	1.61	1.63 (0.27)	1.36	2.17	<0.001
Putamen R	1.48 (0.11)	1.19	1.84	1.81 (0.26)	1.49	2.34	<0.001
Putamen L	1.44 (0.10)	1.22	1.67	1.70 (0.23)	1.37	2.15	0.001
Thalamus R	1.48 (0.13)	1.28	1.87	1.69 (0.11)	1.52	1.93	<0.001
Thalamus L	1.56 (0.13)	1.33	1.96	1.78 (0.15)	1.60	2.07	<0.001
Pons	1.73 (0.16)	1.30	2.02	1.99 (0.20)	1.80	2.41	0.001

SUVR, Standard Uptake Value ratio; PD, Parkinson's disease; Aβ, amyloid-β; SD, standard deviation; R, right; L, left.

Figura 2.2: Tabla 2. SUVR regionales de forbetaben según el estado amiloide.

Considerando toda la cohorte de PD (PDAβ+ y PDAβ-), hubo correlación negativa entre la puntuación MoCA y captación de amiloide florbetaben en las regiones cortical y subcortical. Se observa asociación similar entre la puntuación SDMT y florbetaen SUVR en regiones corticales seleccionadas, las cuales indican que a mayor carga amiloide, peor rendimiento cognitivo.

Es un estudio de PET-amiloide con investigación de la contribución de deposición de amiloide y sus efectos en el rendimiento cognitivo en pacientes con PD no dementes en estadio temprano. Se encuentra presencia de captación de amiloide florbetaben cerebral positivo en el 21 % de la muestra PD al comienzo de la enfermedad y en varias zonas neocorticales y subcorticales.

Se sugiere que la PD por sí sola no confiere un riesgo específico de aumento de la deposición de amiloide. Esta se asocia con aumento de p-tau/Aβ₁₋₄₂ y T-tau/Aβ₁₋₄₂, alineado con un mayor riesgo de patología PD + AD en presencia de T-tau/Aβ₁₋₄₂ y alta carga de amiloide cerebral. Se observa un rendimiento significativamente menor en SDMT en pacientes con PDA±, una medida de la velocidad de procesamiento de la información, que involucra la atención, la memoria de trabajo y los procesos visuales y una correlación significativa con un aumento de los depósitos de amiloide en la cortical.

2.5. Categorical predictive and disease progression modeling in the early stage of Alzheimer's disease [42].

2.5.1. Introducción.

Los modelos de progresión de la enfermedad (DPMs³) dan un diagnóstico mediante el orden explícito de biomarcadores desde estadios normales a patológicos a lo largo del tiempo de forma multivariada. Se basan en análisis longitudinales de cohortes en diferentes etapas.

2.5.2. Materiales.

Se seleccionó el conjunto de datos de la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI⁴) para evaluar el enfoque de los sujetos en dos etapas con diferentes números de visitas. El paquete de información corresponde al ADNI-MERGE R. Nos enfocamos en las siguientes evaluaciones:

- Evaluación de la enfermedad del Alzheimer- Escala cognitiva 11 (*ADAS11, Alzheimer's Disease Assessment*).
- Evaluación de la enfermedad del Alzheimer- Escala cognitiva 13 (*ADAS13*).
- Clasificación de demencia clínica (*CDRSB, Clinical Dementia Rating-Sum of Boxes*).
- Miniexamen del estado mental (*MMSE, Mini-Mental State Examination*).
- Evaluación cognitiva de Montreal (*MoCA, Montreal Cognitive Assessment*).
- Prueba de aprendizaje verbal auditivo inmediato de Rey (*RAVLT, Rey Auditory Verbal Learning Test*).
- Cognición cotidiana (*ECog, Everyday Cognition*).
- Cognición total por participante (*ECogPtTotal, ECog-total by participant*).
- Cognición total por compañero de estudio (*ECogSPTotal, ECog-total by study partner*).
- Cuestionario de evaluación funcional (*FAQ, Functional Assessment Questionnaire*).

Las medidas volumétricas de las imágenes cerebrales se procesaron transversalmente con Freesurfer. Se realizó un segundo análisis entre personas con datos de β -amiloide. Para predecir la conversión de CU a MCI/demencia se eligen pacientes CU y se ve si han convertido a MCI/demencia en el seguimiento mediante un diagnóstico clínico. Los que progresaron a MCI o demencia se denotan como pCU. El tiempo de conversión se establece como el intervalo de tiempo entre el inicio y la primera visita de diagnóstico de MCI/demencia. El grupo CU estable (sCU) está formado por los sujetos CU que no convierten ni a MCI ni demencia. Las últimas

³Disease progression models

⁴Alzheimer's Disease Neuroimaging Initiative

visitas de los sujetos sCU definen sus tiempos de censura.

Para el primer estudio con datos de medidas neuropsicológicas (NM⁵) e imágenes de resonancia magnética (MRI⁶), se dividieron en 316 sujetos sCU y 93 sujetos pCU. Se agregaron 523 pacientes con MCI. En el segundo, agregando marcadores de CSF, se dividieron en 218 sCU y 64 pCU. Se agregaron 399 pacientes con MCI.

2.5.3. Métodos.

Se propone un enfoque en dos etapas para la predicción del diagnóstico categórico y el modelado de progresión de la enfermedad:

- Análisis de supervivencia para determinar subconjuntos de los marcadores multivariados a partir de datos longitudinales para discriminar entre grupos clínicos.
- Los subconjuntos se envían para desarrollar DPMs con enfoques GRACE y LTJMM.

Los marcadores p se medirán desde los n individuos en diferentes tiempos. El modelo LME se expresa:

$$y_{ijk} = x'_{tijk}\beta_k + \alpha_{1ik}t_{ijk} + e_{ijk} \quad (2.1)$$

Donde:

- y_{ijk} : resultado medido k para el individuo i en el instante j .
- t_{ijk} : tiempo de observación a corto plazo.
- x'_{tijk} : fila vector para los efectos fijos.
- β_k : coeficientes de efectos fijos.
- α_{0ik} y α_{1ik} : intersección aleatoria y pendiente específicas del sujeto y del resultado.
- e_{ijk} : error de medición.

Con una población CU, se extraen mediciones longitudinales modelados por LME, estimando para cada sujeto el valor de cada marcador en el tiempo. Se sabía si el sujeto convertía a MCI o demencia en el seguimiento. En los sujetos con pCU, el tiempo de conversión partía desde el inicio del estudio. Para los sCU, se conocía el tiempo de censura.

Se construyó un modelo de Cox extendido para cada tiempo discreto con la siguiente expresión:

⁵ *Neuropsychological measures*

⁶ *Magnetic resonance imaging*

$$h_{ij} = h_j \exp \left(\sum_{k=1}^{p_1} \eta_k \cdot y_{ijk} + \sum_{l=1}^{p_2} \theta_l \cdot z_{il} \right) \quad (2.2)$$

Donde:

- h_j : función de riesgo al inicio del estudio.
- y_{ijk} y z_{il} : variables con el tiempo e independientes.
- η_k y θ_l : coeficientes asociados.

En ambos enfoques se utiliza validación cruzada (CV⁷) anidado para evitar sobreajuste del modelo y estimaciones optimistas del rendimiento del mismo. El procedimiento contiene dos bucles anidados CV:

- Bucle interior: selección de los subconjuntos de características óptimos para los modelos propuestos. Los datos se dividen en datos de entrenamiento y validación.
- Bucle exterior: obtención de una estimación imparcial del rendimiento del modelo. Los datos se divide en modelo y datos de prueba.

En ambos bucles se utiliza un diseño de CV en 10 pliegues. Para cada lazo CV interno se proponen combinaciones de marcadores evaluados en el lazo CV externo.

Los pasos de cada ciclo CV interno son:

1. Método de remuestreo buscando los primeros 10 subconjuntos de cada dimensión con mayor frecuencia en mRMR. Se seleccionaron las 10 combinaciones de características con mayor frecuencia.
2. Los modelos predictivos se construyeron utilizando solo los datos de entrenamiento con los subconjuntos de características candidatas anteriores. Se seleccionaron las 3 combinaciones de marcadores con mejor rendimiento en términos de precisión.

En cuanto a LTJMM, la evolución temporal de los marcadores se expresa como:

$$y_{ijk} = \gamma_k (t_{ijk}^c + \delta_i) + x t_{ijk}^c \beta_k + \alpha_{0ik} + \alpha_{1ik} t_{ijk}^c + e_{ijk} \quad (2.3)$$

Donde:

- γ_k : pendiente específica del resultado con respecto al tiempo desplazado a largo plazo.
- δ_i : cambio de hora que cuantifica la progresión del i-ésimo individuo con respecto a la población que sigue $\delta_i \sim N(0, \sigma_\delta^2)$.

⁷ *Cross-validation*

GRACE permite diferentes formas de curvas monótonas en las trayectorias a largo plazo sin especificación previa. LTJMM impone que las trayectorias a largo plazo deben ser lineales. Previamente al enfoque DPM de los datos, se transforman estos en percentiles para que tengan una escala común. Los marcadores progresan de normal a anormal en una escala vertical común. La escala percentil se normalizó en un rango de 0 (menos grave) a 1 (más grave). Se realizó una transformación percentil ponderada por la desigualdad de los grupos clínicos.

Se asume el cambio de tiempo siguiendo a $\delta_i \sim N(0, \sigma_\delta^2)$. t_{ijk}^c es el año central de las visitas. δ_i es una medida relativa de la progresión de la enfermedad que representa las variabilidades de biomarcadores observadas en la población de entrenamiento.

Para un año inicial de deterioro cognitivo o tiempo cero, t_{onset} :

- Las trayectorias de los sujetos sCU deben estar a la izquierda de t_{onset} . Los marcadores evolucionan en tiempos negativos a largo plazo.
- Las trayectorias de los sujetos pCU deben cruzar t_{onset} hacia valores positivos. Los marcadores evolucionan en tiempos positivos a largo plazo.

2.5.4. Resultados.

Se realizó un análisis multivariante para generar los modelos predictivos propuestos. Se consideran al mismo tiempo:

- 5 medidas MRI basadas en ROI de estructuras corticales y subcorticales.
- 13 NMs.
- 5 biomarcadores de CSF.

Se aplicó un procedimiento de k-fold CV y se construyeron modelos predictivos con sus evaluaciones posteriores. Los modelos predictivos categóricos se seleccionaron teniendo en cuenta las puntuaciones de predicción y el número de veces que se propusieron evaluar sus vectores de características. Las puntuaciones de clasificación fueron similares entre las dos cohortes al igual que las combinaciones de marcadores propuestos para construir los modelos predictivos. Se sugirieron como medidas el volumen hipocampal normalizado (*NHV*, *normalized hippocampal volume*) y las puntuaciones cognitivas y funcionales de ADAS11, FAQ y EcogSPTotal. Se propuso incluir marcadores CSF, pTau o la relación pTau/A β o Tau/A β .

Con el tiempo, las puntuaciones mejoraron en ambas poblaciones especialmente hasta el mes 12. Si se usan la edad, género y años de educación como covariables en el modelado de las trayectorias LME, no se presentan mejoras en las puntuaciones. ADAS11 y FAQ fueron los marcadores que más contribuyeron a la clasificación del deterioro cognitivo. NHV y la edad, la tercera y cuarta, respectivamente.

Tras seleccionar los subconjuntos para cada población, se aplicaron algoritmos DPM (GRACE y LTJMM). Se transformaron las medidas en percentiles. Dado que

las trayectorias a largo plazo de los marcadores fueron casi lineales, se procede a analizar los valores de los marcadores al comienzo de la historia natural o intersecciones y sus pendientes anuales:

- Las intersecciones altas en comparación con el resto de los marcadores muestran que ese marcador es un factor de riesgo en la conversión a deterioro cognitivo.
- Las pendientes muestran la velocidad del marcador en relación con el deterioro cognitivo con el tiempo.

2.6. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data [26].

2.6.1. Introducción.

Se busca en este estudio la ampliación en los modelos de efectos lineales mixtos (*LME, Linear Mixed Effects*) para manejar datos longitudinales de neuroimagen (*LNI, longitudinal neuroimage*) espaciales.

Los modelos longitudinales dedicados como los LME, tienen alguno de los siguientes inconvenientes:

- La selección del modelo se realiza para cada vóxel por separado. Se produce un problema de comparaciones múltiples.
- Los modelos a nivel vóxel no aprovechan la estructura espacial de los datos. Se observan estimadores menos eficientes y con un menor poder estadístico.

Los modelos anteriores que son útiles para el análisis de datos de series temporales, no son adecuados para el análisis de datos LNI por tres razones:

1. Los datos LNI suelen estar desequilibrados. Los puntos de tiempo y el momento de escaneo varía entre sujetos.
2. Se dispone de pocos escaneos longitudinales por sujeto. Se puede obligar a considerar modelos de covarianza más complejos de lo necesario.
3. Algunas suposiciones de modelado para datos de series temporales funcionales no son realistas para datos LNI.

En este estudio se presenta un método para el análisis univariante masivo de datos LNI basado en un modelado de efectos mixtos lineales espaciotemporales (*ST-LME, spatiotemporal LME*). Este enfoque otorga una mejor precisión en la estimación de los parámetros y los grados de libertad estadística.

2.6.2. Material y métodos.

Un enfoque para el análisis univariado de datos LNI es aplicar el LME en cada ubicación espacial (vóxel o vertice de malla) de forma independiente. A este enfoque

se le denomina *LME de vóxeles o vértices (V-LME)*. La idea en LME es permitir que un subconjunto de parámetros de regresión varíe aleatoriamente entre sujetos. La trayectoria media es una combinación de efectos fijos a nivel población y efectos aleatorios a nivel sujeto.

El modelo LME se puede expresar como:

$$Y_i = X_i\beta + Z_ib_i + e_i \quad (2.4)$$

Donde:

- Y_i : vector $n_i \times 1$ de mediciones en serie para el sujeto i .
- n_i : número específico del sujeto de mediciones en serie.
- X_i : matriz $n_i \times p$ de diseño de sujetos para los efectos fijos (género, educación, etc.).
- $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$: vector $p \times 1$ de coeficientes de regresión de efectos fijos desconocidos.
- Z_i : matriz $n_i \times q$ de diseño de sujetos para efectos aleatorios.
- $b_i = (b_{i1}, b_{i2}, \dots, b_{iq})^T$: vector $q \times 1$ de efectos aleatorios.
- $e_i = (e_{i1}, e_{i2}, \dots, e_{in_i})^T$: vector $n_i \times 1$ de errores de medición.

Se establecen los siguientes supuestos:

- $b_i \sim N(0, D)$
- $e_i \sim N(0, \sigma^2 I_{n_i})$

El modelo LME presenta la siguiente media de la población:

$$E(Y_i) = X_i\beta \quad (2.5)$$

La matriz de covarianza temporal no diagonal entre las mediciones en serie del i -ésimo sujeto es:

$$Cov(Y_i) = \Sigma_i = Cov(Z_ib_i) + Cov(e_i) = Z_i D Z_i^T + \sigma^2 I_{n_i} \quad (2.6)$$

Se supone que la estructura de covarianza temporal del modelo LME se comparte entre vóxeles dentro de una región de interés (*ROI, region of interest*). Suponiendo una parcelación de la imagen en regiones homogéneas, se centra la atención en cada una de ellas por separado para su modelado. Y_{ig} se compone de apilar n_i vectores de medias en serie de v_g vóxeles:

$$Y_{ig} = \begin{pmatrix} Y_{ig1} \\ Y_{ig2} \\ \dots \\ Y_{igV_g} \end{pmatrix}$$

Donde:

- g : región que estamos considerando.
- V_g : número de vóxeles o vértices en esta región.
- Y_{ig} : vector ($n_i v_g \times 1$) de medidas para la región g en el sujeto i .
- n_i : número de mediciones en serio de un sujeto en específico.

Se usa la máxima verosimilitud mínima (*REML*, *restricted maximum likelihood*) para estimar los parámetros del modelo asociados con la región g . Se destaca que en el enfoque ST-LME, se realiza una prueba de hipótesis separada en cada vértice. La cantidad de pruebas realizadas y la corrección de comparaciones múltiples es exactamente la misma que un análisis de vértice, como V-LME.

Se supone la existencia de estimaciones aproximadas de los componentes de la covarianza temporal en cada ubicación del cerebro. El objetivo es dividir el dominio de la imagen (R) en n regiones homogéneas (R_1, R_2, \dots, R_r). El algoritmo de segmentación está constituido por dos etapas:

1. Toda la imagen se divide recursivamente en un gran número de pequeñas regiones homogéneas hasta que se cumpla que $H(R_i) = true$, es decir, es homogénea.
2. Estrategia de crecimiento de regiones: las regiones vecinas se fusionan recursivamente si la región resultante aún es homogénea.

Permitimos $H(R_i) = true$ cuando en R_i :

1. Más del 95 % de los vértices de la región tienen una entrada de atributo que es menor que k (similitud entre los componentes de covarianza en una región) desviaciones estándar lejos de la media de la región.
2. La correlación de los mínimos cuadrados ordinarios (*OLS*, *ordinary least squares*) residuales en R_i es mayor que 0.5.

Se proporcionan fórmulas para estimaciones de vértice de los parámetros de covarianza. Estos se basan en estimaciones de mínimos cuadrados ordinarios (*OLS*, *ordinary least squares*) para el LME. Una vez se completa el paso de parcelación, se promedian estimaciones de parámetros dentro de cada región para usarlas como inicio en el REML iterativo.

Hay 4 grupos clínicos en ADNI:

- Controles sanos estables (HC).
- Estables con Deterioro Cognitivo Leve (sMCI).

- MCI convertidor (cMCI).
- Pacientes con AD.

Se analizan los datos de grosor cortical longitudinal en la corteza debido a la demostración de asociación entre la AD y el adelgazamiento cortical.

Se toman dos decisiones importantes en el análisis basado en LME de datos longitudinales:

1. Especificación de variables independientes, variables en el tiempo, que determinan la trayectoria temporal media.
2. Especificación de variables independientes, variables en el tiempo, que determinan la estructura de covarianza.

Para tener en cuenta cualquier posible no linealidad se lleva a cabo un procedimiento de selección de modelos. El modelo final se basa en una trayectoria lineal con dos efectos aleatorios: intercepción y tiempo. En el método ST-LME se aplican cinco iteraciones de maximización de expectativas para mejorar las estimaciones iniciales de vértice de los componentes de covarianza. Se segmenta el cerebro en regiones homogéneas de estimaciones de covarianza similares. Se utiliza la superficie esférica para calcular las distancias en la matriz de correlación espacial. Se centra el estudio en el análisis de la asociación entre la interacción grupo-tiempo y el espesor cortical.

2.6.3. Resultados.

Los sujetos con un solo punto de tiempo fueron descartados del análisis. Las estimaciones de la pendiente se sometieron a un análisis de sujeto cruzados basado en el modelo lineal general (*GLM, General Linear Model*). Las variables independientes de GLM son las mismas que las del efecto fijo del análisis LME (excepción en el tiempo, no incluido). Para los análisis ST-LME, los resultados reflejan la variabilidad en el paso de segmentación.

Se seleccionan aleatoriamente dos conjuntos de muestras AD+HC independientes con el mismo número de sujetos 400 veces. Primero, se evalúa la tasa de error familiar (*FWE, family-wise error*). Los tres métodos proporcionan un control muy bueno de la tasa de error tipo I, siendo V-LME el más conservador. El ST-LME ofrece una sensibilidad y repetibilidad superiores a los métodos de referencia considerados. La diferencia entre el poder estadístico de ST-LME y V-LME tiende a disminuir con el aumento del tamaño de la muestra.

La potencia estadística, la repetibilidad y el control de errores de tipo I que ofrece ST-LME están influenciados por el paso de segmentación y por el tamaño y número de las regiones de segmentación. Se presentan los mapas revelados por los enfoques ST-LME y X-Slope para caracterizar las diferencias de adelgazamiento longitudinal entre los cuatro grupos clínicos (HC, sMCI, cMCI y pacientes con AD.) Las figuras 2.3 y 2.4 muestran los mapas para comparar las tasas de adelgazamiento cortical entre sujetos HC y AD obtenidos por ST-LME y X-Slope. ST-LME revela un grado mucho más amplio de adelgazamiento cortical significativo en comparación con

X-Slope para los análisis MCI a AD frente HC y sMCI frente a cMCI.

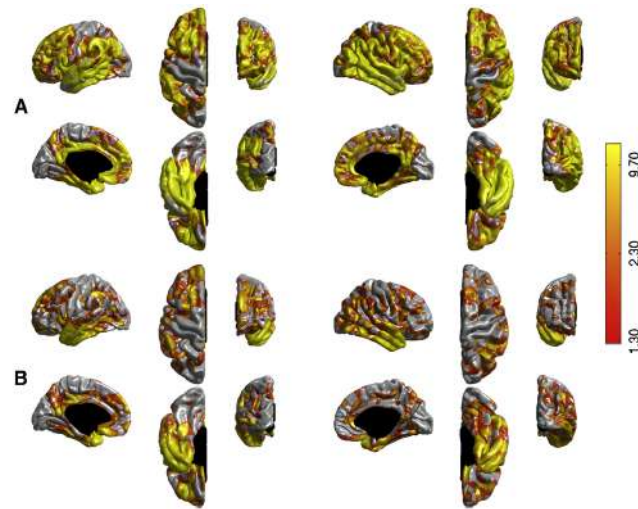


Figura 2.3: Significación estadística no corregida: \log_{10} negativo (p-valor): mapas que comparan las tasas de adelgazamiento cortical longitudinal entre sujetos HC ($N = 210$) y AD ($N = 188$) (de toda la muestra de ADNI) visualizados en la superficie pial de la plantilla FreeSurfer (fsaverage): (A) método ST-LME y (B) método X-Slope. El hemisferio izquierdo se muestra a la izquierda y el hemisferio derecho a la derecha. Los vértices que tienen un p-valor sin corregir inferior a 0.05 se muestran en color. Las filas impares muestran las vistas lateral, superior y anterior. Las filas pares muestran las vistas medial, inferior y posterior. La barra de color muestra el valor de significación correspondiente.

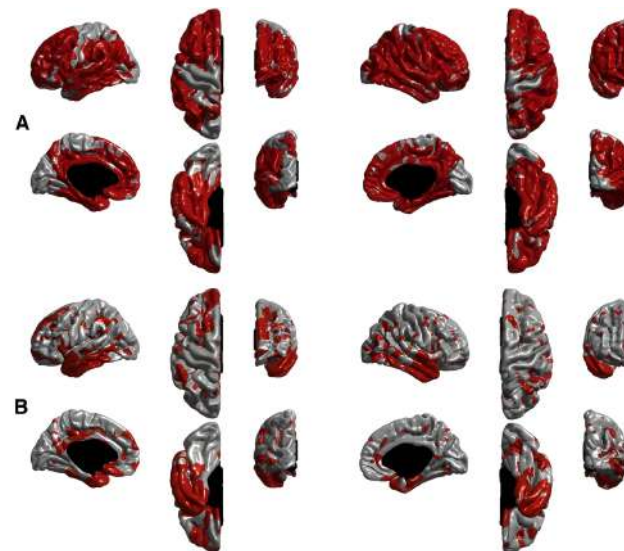


Figura 2.4: Regiones corticales que muestran una diferencia estadísticamente significativa en el adelgazamiento longitudinal entre sujetos HC y AD (en rojo) en toda la muestra de ADNI. Estos mapas se derivaron de umbralizar los valores que se muestran en Figura 2.3 con una corrección FDR en $q = 0,05$. (A) método ST-LME y (B) método X-Slope. ST-LME revela un grado mucho más amplio de adelgazamiento significativo en la AD.

2.7. Estimating long-term multivariate progression from short-term data [34]

2.7.1. Introducción.

El tiempo cero es necesario para ajustarse al modelo estándar de efectos mixtos. La enfermedad de Alzheimer (*AD*, *Alzheimer's Disease*) se desarrolla antes de la aparición de síntomas. ADNI (*Alzheimer's Disease Neuroimaging Initiative*) tiene sujetos diagnosticados como:

- Cognitivamente Normales (CN).
- Deterioro cognitivo leve temprano (EMCI).
- Deterioro cognitivo leve tardío (LMCI).
- AD leve.

Mediante la recopilación de biomarcadores y resultados de evaluaciones cognitivas y funcionales durante años permite que el subconjunto que progresa a AD se use para modelar la progresión de biomarcadores a largo plazo de la enfermedad. En este estudio, se parten de medidas tomadas en un corto periodo de tiempo, en vez de durante años. La regresión de automodelado (*SEMOR*, *Self-MODEling Regression*) ajusta conjuntos de curvas bajo el supuesto de una forma común. Se asume en SEMOR una forma común en toda la población y estiman las curvas del sujeto con las de la población.

2.7.2. Supuestos del modelo.

Se suponen resultados Y_{ij} durante un tiempo t para un individuo i resultado j :

$$Y_{ij} = g_j(t + \gamma_i) + \alpha_{0ij} + \alpha_{1ij}t + \varepsilon_{ij}(t) \quad (2.7)$$

Donde:

- $g_j(t + \gamma_i)$: función monótona diferenciable que traza la progresión a largo plazo.
- t : tiempo.
- γ_i : desfase temporal.
- α_{0ij} : efectos aleatorios independientes del tiempo.
- α_{1ij} : efectos aleatorios dependientes del tiempo.
- $\varepsilon_{ij}(t)$: error residual.

El tiempo de observación a 'corto plazo' se representa con la covariable t . El tiempo de progresión a 'largo plazo' se representa con $(t + \gamma_i)$. γ_i es el cambio de tiempo específico del sujeto desconocido.

2.7.3. El Algoritmo.

Los parámetros desconocidos (g_j , γ_i y α) se estiman utilizando estimaciones de los demás parámetros con un algoritmo de suma de cuadrados residual (RSS , *Residual Sum of Squares*). Los residuos parciales se estiman con el vector $(R_{ij}^g(t), R_{ij}^\gamma(t), R_{ij}^\alpha(t))$.

Algoritmo 1 Algoritmo de GRACE

Entrada: Definición de cálculo de residuos parciales

Salida: Estimación de los parámetros g_i , γ_i y α

- 1: Inicializar $\gamma_i=0$
 - 2: Igualar parámetros $\alpha=0$
 - 3: **mientras** no converjan $R_{ij}^g(t), R_{ij}^\gamma(t)$ **hacer**
 - 4: Estimar g_j con el suavizado del residuo parcial $R_{ij}^g(t)$
 - 5: Estimar α del LME de $R_{ij}^\alpha(t)$
 - 6: Comprobar convergencia de $R_{ij}^g(t), R_{ij}^\alpha(t)$
 - 7: **fin mientras**
 - 8: **mientras** no converjan $R_{ij}^g(t), R_{ij}^\gamma(t), R_{ij}^\alpha(t)$ **hacer**
 - 9: **mientras** no converjan $R_{ij}^g(t), R_{ij}^\alpha(t)$ **hacer**
 - 10: Estimar g_j con el suavizado del residuo parcial $R_{ij}^g(t)$
 - 11: Estimar α del LME de $R_{ij}^\alpha(t)$
 - 12: Comprobar convergencia de $R_{ij}^g(t), R_{ij}^\alpha(t)$
 - 13: **fin mientras**
 - 14: Tomar estimación de g_j
 - 15: Igualar a cero los parámetros α y $\varepsilon_{ij}(t)$
 - 16: Estimar $R_{ij}^\gamma(t)$ en todo instante.
 - 17: Comprobar convergencia de los tres residuos.
 - 18: **fin mientras**
 - 19: Asumir g_i, γ_i, α como resultado de estimar los residuos.
-

2.7.4. Resultados y discusión.

Primero se realiza una conversión a percentiles para introducirlas en el algoritmo y homogenizar los resultados debido a las diferentes escalas entre las diferentes medidas de los sujetos con una función de distribución acumulación empírica (*ECDF*, *Empirical aCumulative Distribution Function*). El 0 representa que el valor es menos severo que el 100, el máximo.

La figura 2.5 muestra en el lado izquierdo las trayectorias para la población amiloide (A+). En el lado derecho se muestra las de la población con APOE ϵ 4. Son curvas monótonas y flexibles con alto grado de linealidad.

El "tiempo cero"⁸ de transición a demencia emplea medidas de la Mini Prueba del Estado Mental (MMSE⁹). Así se puede situar el tiempo cero de las trayectorias en la transición a demencia AD.

⁸También denominado *umbral*.

⁹Mini Mental State Examination

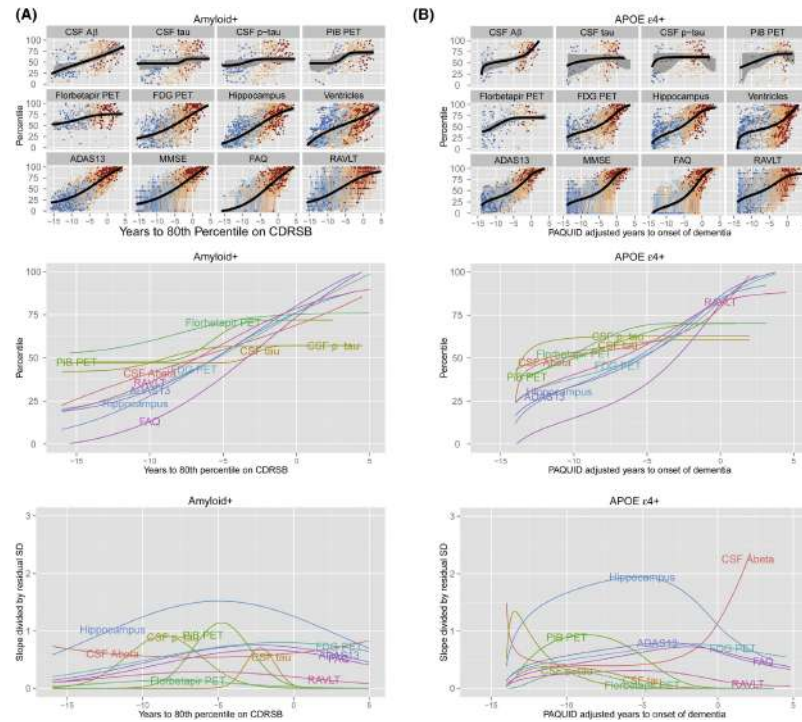


Figura 2.5: (A) Sujetos A+ de ADNIA. Portadores de APOE ϵ 4 en ADNI. Los paneles superiores muestran cada una de las trayectorias medias superpuestas sobre las observaciones a nivel de sujeto de 579 A+ y 570 APOE ϵ 4 individuos, coloreados por diagnóstico. Los colores en el panel superior representan el diagnóstico en la línea de base de ADNI: cognitivamente normal (CN) en azul oscuro, deterioro cognitivo leve temprano (EMCI) en azul claro, deterioro cognitivo leve tardío (LMCI) en rojo claro y enfermedad de Alzheimer (AD) en oscuro rojo. Las regiones sombreadas en gris, donde están visibles en los paneles superiores, representan bandas de confianza del 95% de arranque. Los paneles centrales muestran todas las trayectorias a la vez. A la izquierda, el tiempo se ha desplazado de modo que el tiempo cero representa el tiempo en el que la trayectoria media de la suma de cajas de la escala de calificación clínica de demencia (*CDRSB*, *Clinical Dementia Rating Scale Sum of Boxes*) (no se muestra) alcanza el percentil 80. A la derecha, el tiempo se ha ajustado utilizando trayectorias de miniexamen del estado mental (PAQUID) a largo plazo, de modo que el tiempo cero representa el tiempo estimado hasta el inicio de la demencia. Los paneles inferiores muestran tasas de cambio estandarizadas por desviación estándar residual (SD). CSF, líquido cefalorraquídeo; A β , amiloide- β ; p-tau, tau fosforilada; PiB, compuesto B de Pittsburgh; PET, tomografía por emisión de positrones; FDG, fluorodesoxiglucosa; ADAS13, la Escala de Evaluación de la Enfermedad de Alzheimer de 13 ítems–Subescala Cognitiva; MMSE, Mini-Examen del Estado Mental; FAQ, Cuestionario de Actividades Funcionales del Estudio Cooperativo de la Enfermedad de Alzheimer; RAVLT, Test de Aprendizaje Auditivo Visual de Rey; CN, cognitivamente normal; LMCI, deterioro cognitivo leve tardío; AD, enfermedad de Alzheimer.

2.8. Conclusión del capítulo

Tras el análisis de los diferentes documentos que han sido de utilidad en el estudio, se presentan descripciones de cada uno, de forma breve, sobre los puntos de mayor interés que han sido aportados al trabajo.

- **Aarsland et al. (2021)** [7]

Los autores estudian los factores de riesgo en la PD. La atención de los factores de riesgo se ha centrado en los cambios cognitivos tempranos, alucinaciones visuales, edad avanzada y cambios en biomarcadores. Además, analizan la pro-

gresión que se produce desde sujetos PD a PD-MCI y PDD. Por último, se exponen tratamientos del MCI, tanto desde un enfoque tanto no farmacológico como farmacológico.

- **Jack et al. (2018)** [6]

Este estudio se realiza para el diagnóstico de la AD con biomarcadores en personas vivas. Estos biomarcadores se agrupan en los depósitos de β -amiloide, pTau y neurodegeneración, AT(N). La clasificación ATN agrupa biomarcadores de imágenes y biofluidos. Además, la consideración de AD como constructo biológico permite la caracterización y comprensión más precisa de los eventos que conducen al deterioro cognitivo asociado a la enfermedad.

- **Schrag et al. (2016)** [2]

El artículo realizado por estos autores proporciona conocimiento en la realización de modelos predictivos en el diagnóstico de la PD. Ofrecen un punto de partida de biomarcadores (imágenes DAT y del CSF) y procedimientos estadísticos para futuros estudios e investigaciones. Estudia la variable de forma longitudinal. A esto, se le suma la importancia de utilizar modelos multivariantes, superando a los univariantes, y que pueden combinar datos de distinto tipo, ofreciendo mejores resultados. En este estudio se establecen como predictores más importantes, la edad, gravedad motora y depresión.

- **Fiorenzato et al. (2018)** [16]

Este documento estudia la relación de la PD con la patología amiloide (A+) debido a la concentración de $A\beta$ en la corteza cerebral. Los resultados de este estudio no relacionan el riesgo de la enfermedad con el aumento de amiloide. En cambio, la combinación de amiloide y α -sinucleína dan lugar a un biomarcador capaz de predecir el deterioro cognitivo de forma longitudinal. Se enfatiza en la importancia de los primeros años de la enfermedad para el futuro de la misma.

- **Platero et al. (2021)** [42]

Este artículo se introduce en el análisis de supervivencia combinado con LME para mejorar la realización de modelos predictivos que diagnostiquen la enfermedad como primer enfoque. A esto, se le añade un segundo enfoque con la comparación de modelos predictivos y clasificación de sujetos (convierten o no a MCI) que hacen añadir unas guías para incluir características clínicas (medidas MRI, pruebas neuromotoras) que dan modelos equilibrados de forma longitudinal. Se obtienen mejores resultados en el primer enfoque debido a la no homogeneidad de los grupos.

- **Bernal et al. (2013)** [26]

Este estudio y el anterior [27] se encargan de presentar la realización de la secuencia en un análisis masivo univariante en pacientes con AD. En este análisis, se caracterizan los diferentes grupos en función del volumen y espesor cortical en diferentes ROI. Los resultados obtenidos sirven de referencia de control de

la calidad de la población.

- **Donohue et al. (2014)** [34]

Los autores de este trabajo sirve como manual para la realización de modelos de progresión de la PD con el paquete GRACE propuesto por el autor a lo largo del tiempo con trayectorias *long-term*. Nos permite visualizar diferentes trayectorias para cada biomarcador del estudio, observando el modelado de la enfermedad. Este estudio, nos permite entender la interpretación de los resultados a partir del *script* ejecutado en R y que se puede ver en el capítulo 4.

Capítulo 3

Datos empleados, herramientas y materiales

3.1. Introducción.

En el capítulo se presentan los datos, herramientas y materiales necesarios para este estudio. Para ello, se muestra tanto la base de datos establecida para la extracción de la información como el software necesario para su procesamiento.

3.2. Datos empleados.

En este estudio se han empleado una población con pacientes incluidos en la base de datos PPMI. Esta población tiene tres tipos de datos:

1. Datos clínicos.
2. Datos de test motores.
3. Datos provenientes de medidas PPMI.

3.2.1. Base de datos. PPMI.

La iniciativa PPMI (*Parkinson's Progression Markers Initiative*¹) es un grupo clínico cuya función es principalmente la de recopilar y analizar grupos y biomarcadores para posibles estudios y avances en el área de la PD [43] (Figura 3.1).



Figura 3.1: Logotipo de la base de datos PPMI.

¹Iniciativa de Marcadores de Progresión de Parkinson

PPMI consigue realizar su adquisición de datos y la elaboración de estudios en varios países alrededor del mundo. En estos países², los centros clínicos participantes en PPMI están especializados en tratamiento e investigación de enfermedades neurológicas.

Los métodos para la extracción de datos utilizados con mayor frecuencia en PPMI son:

- **Imágenes MRI:** Ponderación de las imágenes en T1 y T2.
- **Muestras biológicas:** Medidas de marcadores neurológicos como ratios (ej. $A\beta/T\tau$) o diferentes niveles (ej.: α -sinucleína).
- **Medidas clínicas:** Pruebas realizadas en las diferentes visitas a los pacientes. Se realizan por ejemplo: MoCA, UPDRS o test de temblor asimétrico. También se incluyen otras variables de la vida del paciente (sexo, etnia, edad, años de educación, etc.).

PPMI tiene centros con autorización y verificación mediante protocolos para que se trate de sujetos anónimos con los datos que se incluyen en la base. Los centros de investigación y universidades pueden acceder de forma sencilla y segura a esta base de datos con el carácter anónimo de los sujetos. Para acceder es necesario solicitar la creación de una cuenta a PPMI en la que se especifica el objetivo del estudio, la institución que hace uso de la base de datos y el tutor responsable del trabajo. Cuando la solicitud ha sido aceptada, se puede acceder a la base datos.

En este estudio, se parten de 423 sujetos PD y 196 sujetos control (HC^3). De estos 423 sujetos PD, se establece la distribución que se muestra en la tabla 3.1.

PD-CU	PD-MCI	PD-Dementia
323	83	17

Tabla 3.1: Distribución de los sujetos PD. PD-CU: Sujetos PD sin deterioro cognitivo. PD-MCI: Sujetos PD que desarrollan deterioro cognitivo leve. PD-Dementia: Sujetos PD que desarrollan demencia.

3.3. Herramientas.

3.3.1. Matlab.

Matlab⁴ es un entorno utilizado en una gran cantidad de algoritmos para el trabajo que se realiza actualmente en diferentes áreas como ingeniería o biomedicina entre otras (figura 3.2) [31].

Matlab tiene IDE⁵. Además, Matlab utiliza lenguaje de programación M, específico de este programa. Aún así, es compatible con otros programas con sus propios lenguajes (C, C++, Java, Python, etc.). Las funciones principales de Matlab son:

²Descargas mayoritarias en Estados Unidos, Canadá, Reino Unido, China o Italia.

³Healthy Control

⁴Abreviatura de MATrix LABoratory.

⁵Entorno de desarrollo integrado.



Figura 3.2: Logotipo del software Matlab.

- Manipulación de matrices.
- Desarrollo de algoritmos.
- Comunicación con otros programas y dispositivos Hardware.
- Representación de datos y algoritmos.
- Creación de interfaces de usuario (GUI).

En nuestro caso, lo hemos utilizado para:

- Organización de datos en tablas.
- Relaciones entre poblaciones de sujetos.
- Realización de resultado gráficos.
- Modelos predictivos.
- Modelos de progresión.

3.3.2. Filezilla.

Filezilla es un software FTP⁶ de licencia libre y código abierto en el que se parte de un cliente y un servidor (Figura 3.3) [20]. Se permite mediante este programa la subida y descargas a nuestra máquina (servidor) de los distintos documentos, imágenes o scripts anteriores o desarrollados por el usuario.



Figura 3.3: Logotipo del software Filezilla.

⁶File transfer protocol.

3.3.3. PuTTY.

PuTTY⁷ es un cliente para protocolos SSH⁸, Telnet, rlogin y TCP raw con software libre para la realización de conexión remota (Figura 3.4) [47]. En este estudio se utiliza para la conexión remota con la máquina de sistema operativo Linux y utilización del software que esta posee.



Figura 3.4: Logotipo del software PuTTY.

3.3.4. Xming.

Xming es un servidor que muestra una interfaz gráfica a través de ventanas en la realización de una conexión remota (Figura 3.5) [23]. Es usado en protocolos SSH con sesiones X11 de forma segura con otras máquinas. Se utiliza junto con PuTTY en la conexión con la máquina para poder visualizar programas que puede ejecutar.



Figura 3.5: Logotipo del software Xming.

3.3.5. LaTeX.

LaTeX es un software que permite componer textos generalmente planteados para que presenten una tipografía de alta calidad (Figura 3.6) [28]. Este programa, a su vez, se compone por gran cantidad de macros de comandos denominados TeX, lenguaje de composición tipográfica⁹.

⁷Port Unique TeleTYpe

⁸Secure Shell

⁹Ideado por Leslie Lamport (1986) y creado por Donald Knuth (1978).



Figura 3.6: Logotipo del software LaTeX.

LaTeX es un software relativamente sencillo y con código abierto. Esto hace que se creen bibliotecas adecuadas a inclusión de otros elementos no textuales (ecuaciones, imágenes, referencias, etc.).

3.3.6. Procesamiento en CeSViMa. Magerit-3.

El CeSViMa¹⁰ es un centro de la UPM¹¹ (Figura 3.7) [4]. Este centro ofrece prestaciones potentes de computación en el ámbito científico y empresarial.



Figura 3.7: Logotipo del software CeSViMa.

Magerit-3 es un supercomputador del CeSViMa. Se utiliza para aplicaciones de altas prestaciones (HPC¹²). Magerit cuenta con 68 nodos, cada uno con:

- Un procesador Intel Xeon Gold 6230¹³.
- 192GB memoria RAM.
- 480GB memoria en disco SSD.

Se hace uso de este recurso para realizar operaciones que serían largas (varias semanas) y complejas con un elevado número de recursos no disponibles en las máquinas de la ETSIDI¹⁴.

Se ha incluido el apéndice C para una mayor profundidad en el uso de Magerit para usuarios que no lo han utilizado y les sirva de guía en sus primeros pasos.

¹⁰Centro de Supercomputación y Visualización de Madrid

¹¹Universidad Politécnica de Madrid

¹²High Performance Computing

¹³20 cores @ 2.1 GHz y 40 hilos.

¹⁴Escuela Técnica Superior de Ingeniería y Diseño Industrial.

Capítulo 4

Métodos.

4.1. Introducción.

En este capítulo se recogen los métodos usados para el procesado de los sujetos utilizados en este estudio para alcanzar objetivos y el flujo de desarrollo. En este trabajo, se sigue la línea de otros trabajos [48, 1].

Primero, se expone la determinación del deterioro cognitivo leve. A continuación, se definen los tiempos de conversión y censura utilizados en los análisis de supervivencia. Este capítulo dedica otra parte de organización de datos al inicio del estudio en características y grupos clínicos para su posterior análisis de resultados. También una segunda parte de desarrollo y análisis de modelos predictivos para la progresión de los sujetos.

Todos los algoritmos desarrollados se encuentran en el usuario *jpalomino* de la siguiente dirección IP:

138.100.100.132:22

4.2. Determinación deterioro cognitivo leve (MCI).

La evolución del carácter cognitivo de los sujetos PD se basa en la Sociedad de Trastornos del Movimiento [48] según los siguientes factores y que se categoriza cada 12 meses:

- Queja cognitiva del paciente o del informante.
- Deterioro cognitivo establecido por un número menor o igual a 2 puntuaciones en pruebas por debajo de la media estandarizada.
- No se presenta deterioro funcional por el deterioro cognitivo.

4.3. Tiempos de conversión y censura.

En el análisis y estudio de un paciente en su conversión o no a MCI o demencia durante el estudio, se realiza un denominado *análisis de supervivencia*. Se determinan 3 grupos:

- PD-CU: Sujetos que no convierten a MCI.
- PD-MCI: Si en la última evaluación cognitiva el resultado de MCI es positivo en el sujeto PD.
- PDD: Si en la última evaluación cognitiva el resultado de demencia es positivo en el sujeto PD.

Además, en este análisis de supervivencia se establecen dos tiempos de interés para el entendimiento de este estudio:

- **Tiempo de conversión:** Se establece como la primera visita en la que el paciente PD tiene una evaluación positiva de MCI o PDD. Por lo tanto, está establecido por los sujetos PD-MCI y PDD.
- **Tiempo de censura:** Se establece como la última vista de los pacientes PD-CU. Por lo tanto, está establecido por los sujetos PD-CU.

Por último, estas evaluaciones se han desarrollado según dos criterios de evaluación cognitiva:

- ***cogstate***¹: Diagnóstico del investigador del sitio del estado cognitivo (se establece MCI cuando es igual a 2 y PDD cuando es igual a 3).
- ***MCI_testscores***: Se basa en puntuaciones de pruebas cognitivas (2 o más puntuaciones > 1.5 SD por debajo de la media estandarizada). Se establece MCI o demencia cuando es igual a 1.

4.4. Análisis al inicio del estudio.

Para comenzar con el estudio de la población que se posee, se realiza un análisis transversal al inicio del estudio que nos permita diferenciar en primer lugar entre los grupo control (HC) y los que poseen la enfermedad de Parkinson (PD).

En segundo lugar, se diferencia los sujetos con PD entre sujetos con patología amiloide (PD-A+) y sin ella (PD-A-). Después, se separan a los sujetos PD según su estado cognitivo en PD-CU, PD-MCI y PDD.

Con estas diferencias de grupos y sus análisis transversales al inicio podemos establecer factores y predictores de riesgo en la enfermedad de Parkinson.

Por último, se analiza el perfil AT como constructos biológicos que caracteriza a los sujetos con la enfermedad de Alzheimer. Este perfil es analizado en los sujetos PD para establecer la conexión entre el declive cognitivo de la PD y dichos constructos de la AD.

¹También denominado en el documento como SITE.

4.4.1. Análisis al inicio del estudio de sujetos PD y HC

Para realizar el análisis al inicio del estudio (*baseline*) entre sujetos PD y HC se utiliza el *script* que se encuentra en */demo/Baseline_PPMI/PDvsHC_bl.m*.

Este algoritmo (2) nos permite realizar un análisis de los diferentes biomarcadores entre ambos grupos. Este algoritmo utiliza *getDataPPMI_HCvsPD_schrag16_v1.m* y *summary_bsl_PDvsHC.m* para la separación de grupos y el muestreo de características seleccionadas. Se pueden encontrar estos códigos fuente en la carpeta */demo/Baseline_PPMI/selected_aux/*

Algoritmo 2 PDvsHC.bl

Entrada: Población y tipo de características

Salida: Medidas de los grupos PD y HC

```

1: Configuración de los tipos de características a obtener y la población
2: Diferenciar entre población PD y HC
3: Máscara para datos solo del baseline
4: Categorizar columna APPRDX en PD o HC en función de si aparece un 1 o un 2
5: Introducción de características, categorización, nombre de características y PPMI filtrados al
  baseline
6: Dar un número a cada categorización y contar el número de diferentes grupos
7: si número de grupos >3 entonces
8:   número de grupos=3
9: fin si
10: si número de grupos=3 entonces
11:   número de grupos=2
12: fin si
13: si número de grupos=2 entonces
14:   p_test=TRUE
15: fin si
16: para i=1:número de grupos hacer
17:   Máscara del grupo i
18:   Filtrar grupo i
19:   Mostrar número de sujetos, resultado de mujeres y APOE4
20: fin para
21: para i=1:número de características hacer
22:   para j=1:número de grupo hacer
23:     Mostrar resultados de la característica i del grupo j
24:   fin para
25:   si p_test==TRUE entonces
26:     Calcular p-valor entre los grupos
27:   fin si
28:   si característica=MOCA,ABETA,ASYN,TAU,PTAU entonces
29:     Mostrar diagramas cajas-bigotes
30:   fin si
31: fin para
32: Mostrar relación ABETAvsASYN, TAUvsPTAU y PTAUvsASYN y valor de correlación

```

4.4.2. Análisis al inicio del estudio de sujetos PD-A+, PD-A-, HC-A+ y HC-A-

Tras la separación de los sujetos PD y HC, para realizar el análisis *baseline* entre sujetos con o sin patología amiloide (A+ y A-, respectivamente) se utiliza el *script*

que se encuentra en `/demo/Baseline_PPMI/PDvsHC_amiloide_bl.m`.

Este algoritmo (3 y 4) nos permite realizar un análisis de los diferentes biomarcadores entre estos grupos, igual que en el caso anterior, añadiendo la separación debido al factor amiloide. Este algoritmo utiliza `getDataPPMI_amiloide_PD_schrag16_v1.m`, `summary_bsl_PD_amiloide.m`, `getDataPPMI_amiloide_HC_schrag16_v1.m`, `summary_bsl_HC_amiloide.m` y `amyloid_groups.m` para la separación de grupos y el muestreo de características seleccionadas. Estos *scripts* están ubicados en la carpeta `/demo/Baseline_PPMI/selected_aux/`.

4.4.3. Análisis al inicio del estudio de sujetos PD-CU, PD-MCI y PDD

Tras la separación de los sujetos PD, para realizar el análisis *baseline* del estado cognitivo, se separa entre sujetos CU, MCI y con demencia. Se utiliza el *script* que se encuentra en `/demo/Baseline_PPMI/PD_MCI_Dementia_bl_anova.m`. Este algoritmo (5) nos permite realizar un análisis de los diferentes biomarcadores entre los tres grupos, igual que en los casos anteriores.

A esto se le añade el análisis anova y la comparación múltiple con el método de Bonferroni (se explica con mayor profundidad en el apéndice B). Este algoritmo utiliza `getDataPPMI_MCI_schrag16_v1.m` y `summary_bsl_MCI_Dementia_v2.m` para la separación de grupos y el muestreo de características seleccionadas. Están ubicados en la carpeta `/demo/Baseline_PPMI/selected_aux/`.

Después de obtener los resultados de las diferentes características en los distintos grupos según su estado cognitivo, se procede a la obtención de los tiempos de censura y de conversión con la visualización de los histogramas en el análisis de supervivencia. Para ello se ha utilizado el *script* que podemos encontrar en `/demo/Baseline_PPMI/histogram_MCI_and_dementia.m`. Este algoritmo (6) permite visualizar 3 histogramas diferentes:

1. Histograma de conversión de PD-CU a PD-MCI o PD-Demencia.
2. Histograma de conversión, separando de PD-CU a PD-MCI y de PD-CU a PD-Demencia.
3. Histograma con los tiempos de censura (PD-CU que no convierten).

Este código fuente utiliza también los *scripts*: `getDataPPMI_MCI_schrag16_v1.m`, `summary_year_MCI_and_dementia.m` y `summary_year_MCI_or_dementia.m`. Al igual que en los casos anteriores, se puede encontrar en `/demo/Baseline_PPMI/selected_aux/`.

4.4.4. Predictores y factores de riesgo

Tras conocer las discrepancias de las distintas medidas y biomarcadores entre los diferentes grupos según su estado cognitivo, se puede establecer un grupo de factores de riesgo a analizar en nuestra población y confirmar que depende de esos valores que se encuentre en un grupo u otro. Para ello, se ha realizado el *script* que

Algoritmo 3 PDvsHC_amiloide_bl

Entrada: Población y tipo de características**Salida:** Medidas de los grupos PD-A+, PD-A-, HC-A+ y HC-A-

```

1: Configuración de los tipos de características a obtener y la población
2: Obtener población PD y diferenciar entre población A+ y A- si  $A\beta < 683.45$ 
3: Máscara para datos solo del baseline
4: Categorizar columna amyloid_positive en PD-A+ o PD-A- en función de si aparece un 0 o un
  1
5: Introducción de características, categorización, nombre de características y PPMI filtrados al
  baseline
6: Dar un número a cada categorización y contar el número de diferentes grupos
7: si número de grupos >3 entonces
8:   número de grupos=3
9: fin si
10: si número de grupos=3 entonces
11:   número de grupos=2
12: fin si
13: si número de grupos=2 entonces
14:   p_test=TRUE
15: fin si
16: para i=1:número de grupos hacer
17:   Máscara del grupo i
18:   Filtrar grupo i
19:   Mostrar número de sujetos, resultado de mujeres y APOE4
20: fin para
21: para i=1:número de características hacer
22:   para j=1:número de grupo hacer
23:     Mostrar resultados de la característica i del grupo j
24:   fin para
25:   si p_test==TRUE entonces
26:     Calcular p-valor entre los grupos
27:   fin si
28:   si característica=MOCA,ABETA,ASYN,TAU,PTAU entonces
29:     Mostrar diagramas cajas-bigotes
30:   fin si
31: fin para
32: Mostrar relación ABETAvsASYN, TAUvsPTAU y PTAUvsASYN
33: Configuración de los tipos de características a obtener y la población
34: Obtener población HC y diferenciar entre población A+ y A- si  $A\beta < 683.45$ 
35: Máscara para datos solo del baseline
36: Categorizar columna amyloid_positive en HC-A+ o HC-A- en función de si aparece un 0 o un
  1
37: Introducción de características, categorización, nombre de características y PPMI filtrados al
  baseline
38: Dar un número a cada categorización y contar el número de diferentes grupos
39: si número de grupos >3 entonces
40:   número de grupos=3
41: fin si
42: si número de grupos=3 entonces
43:   número de grupos=2
44: fin si
45: si número de grupos=2 entonces
46:   p_test=TRUE
47: fin si

```

Algoritmo 4 PDvsHC.amiloide_bl (Continuación)

```

1: para i=1:número de grupos hacer
2:   Máscara del grupo i
3:   Filtrar grupo i
4:   Mostrar número de sujetos, resultado de mujeres y APOE4
5: fin para
6: para i=1:número de características hacer
7:   para j=1:número de grupo hacer
8:     Mostrar resultados de la característica i del grupo j
9:   fin para
10:  si p_test==TRUE entonces
11:    Calcular p-valor entre los grupos
12:  fin si
13:  si característica=MOCA,ABETA,ASYN,TAU,PTAU entonces
14:    Mostrar diagramas cajas–bigotes
15:  fin si
16: fin para
17: Mostrar relación ABETAvsASYN, TAUvsPTAU y PTAUvsASYN

```

se encuentra en */demo/Baseline_PPMI/PD_risk_factors.m*.

Con este algoritmo (7), se puede separar entre los tres grupos cognitivos y analizar el valor de los factores y predictores. Para utilizar este código fuente, se apoya en otros *scripts*: *getDataPPMI_MCI_schrag16_v1.m*, *summary_risk_f.m*. Estos códigos auxiliares se encuentran en el directorio */demo/Baseline_PPMI/selected_aux/*.

4.4.5. Perfil AT en sujetos PD

Tras el análisis de las distintas medidas en los diferentes grupos clínicos y cognitivos de los sujetos que poseen la enfermedad de Parkinson, se han establecido una serie de características que podrían actuar como factores de riesgo y predictores de la enfermedad.

Otro de los objetivos de este estudio consiste en establecer esa relación del declive cognitivo en sujetos con PD con los constructos biológicos de la AD. Para ello, se enfoca el análisis en el perfil AT que se refleja en el artículo de Jack et al. [6], procedente de la $A\beta$ y de la pTau.

Se establece que un sujeto es A+, cuando este tiene un $A\beta < 683.45$ pg/ml. Además se establece que un sujeto es perfil T+ cuando este tiene un pTau > 23 pg/ml. Si un sujeto tiene A+, se denomina Alzheimer continuo. Para ello, se ha realizado un código fuente que se puede encontrar en */demo/Baseline_PPMI/PD_ATNperfiles.m*. Este *script* hace uso de *getDataPPMI_MCI_schrag16_v1.m*, *summary_ATN_baseline.m* y *summary_ATN_long.m*. Según el nombre que reciben estos *scripts*, se observa que el algoritmo (8) analiza estos perfiles tanto al inicio del estudio como de forma longitudinal.

En este análisis longitudinal, en el caso del perfil A+, se utiliza el mínimo valor de $A\beta$ en el tiempo. En el caso del perfil T+, se utiliza el máximo valor de pTau en

Algoritmo 5 PD_MCI_Dementia_bl_anova

Entrada:**Salida:** Medidas de los grupos PD-CU, PD-MCI y PDD

- 1: Configuración de los tipos de características a obtener, población y criterio MCI
 - 2: Obtener población PD y los diferentes criterios MCI calculados
 - 3: Calcular tiempo de conversión
 - 4: Máscara para datos solo del baseline
 - 5: Categorizar columna *cogstate_end* en PD-CU, PD-MCI o PDD en función de si aparece un 1, 2 o 3
 - 6: Introducción de características, categorización, nombre de características y PPMI filtrados al baseline
 - 7: Dar un número a cada categorización y contar el número de diferentes grupos
 - 8: **si** número de grupos >3 **entonces**
 - 9: número de grupos=3
 - 10: **fin si**
 - 11: **si** número de grupos=3 **entonces**
 - 12: número de grupos=3
 - 13: multi_comp=TRUE
 - 14: p_test=FALSE
 - 15: **fin si**
 - 16: **si** número de grupos=2 **entonces**
 - 17: p_test=TRUE
 - 18: multi_comp=FALSE
 - 19: **si no**
 - 20: p_test=FALSE
 - 21: **fin si**
 - 22: **para** i=1:número de grupos **hacer**
 - 23: Máscara del grupo i
 - 24: Filtrar grupo i
 - 25: Mostrar número de sujetos, resultado de mujeres y APOE4
 - 26: **fin para**
 - 27: **para** i=1:número de características **hacer**
 - 28: **para** j=1:número de grupo **hacer**
 - 29: Mostrar resultados de la característica i del grupo j
 - 30: **fin para**
 - 31: **si** multi_comp==TRUE **entonces**
 - 32: Realizar multicomparación entre los grupos para conocer discrepancias
 - 33: **fin si**
 - 34: **fin para**
 - 35: Mostrar relación ABETAvsASYN, TAUvsPTAU y PTAUvsASYN
-

Algoritmo 6 histogram_MCI_and_dementia

Entrada:**Salida:** Histogramas de los tiempos de conversión y censura

- 1: Configuración de los tipos de características a obtener, población y criterio MCI
 - 2: Obtener población PD y los diferentes criterios MCI calculados
 - 3: Calcular tiempo de conversión
 - 4: Máscara para datos solo del baseline
 - 5: Aplicar máscara del baseline a PPMI
 - 6: Máscara para datos con MCI=1
 - 7: Máscara para datos con MCI=0
 - 8: Categorizar columna *convertTime* en 0,1,2,3,4 o 5 en función de el año de conversión
 - 9: Introducción de categorización, PPMI filtrados al MCI=1 y PPMI filtrados al MCI=0
 - 10: Dar un número a cada categorización y contar el número de diferentes grupos
 - 11: **para** i=1:número de grupos **hacer**
 - 12: Máscara del grupo i
 - 13: Contar el número de sujetos con esa máscara
 - 14: **fin para**
 - 15: Mostrar histograma de la población con MCI=1 según el año de conversión
 - 16: Mostrar histograma de la población con MCI=0 según el año de censura
 - 17: Introducción de PPMI filtrados al MCI=1
 - 18: Categorizar columna *cogstate_end* en MCI o Dementia en función de si el valor es 2 o 3
 - 19: Máscara para el grupo que convierte a MCI
 - 20: Aplicar máscara para el grupo que convierte a MCI en la población MCI=1
 - 21: Máscara para el grupo que convierte a Dementia
 - 22: Aplicar máscara para el grupo que convierte a Dementia en la población MCI=1
 - 23: Mostrar histograma de la población separando la población que convierte a MCI según el año de conversión
-

Algoritmo 7 PD_risk_factors

Entrada:**Salida:** Valores de los factores y predictores de riesgo para los grupos cognitivos

- 1: Configuración de los tipos de características a obtener, población y criterio MCI
 - 2: Obtener población PD y los diferentes criterios MCI calculados
 - 3: Calcular tiempo de conversión
 - 4: Máscara para datos solo del baseline
 - 5: Aplicar máscara del baseline a PPMI
 - 6: Máscara para datos con MCI=1
 - 7: Máscara para datos con CU=1
 - 8: Máscara para datos con Dementia=1
 - 9: Introducción de PPMI filtrados al MCI=1, PPMI filtrados al CU=1 y PPMI filtrados al Dementia=1.
 - 10: **para** cada grupo clínico **hacer**
 - 11: **para** cada factor/predictor (MoCA, APOE4, A+, Género, REM, UPSIT, Tau) **hacer**
 - 12: Filtrar los que poseen ese factor/predictor
 - 13: Calcular cantidad del grupo con ese factor/predictor
 - 14: **fin para**
 - 15: Mostrar resultados de los factores/predictores de riesgo del grupo
 - 16: **fin para**
-

los años de estudio.

Algoritmo 8 PD_ATNperfiles

Entrada:

Salida: Valores de los perfiles en los grupos cognitivos

- 1: Configuración de los tipos de características a obtener, población y criterio MCI
 - 2: Obtener población PD y los diferentes criterios MCI calculados
 - 3: Calcular tiempo de conversión
 - 4: Máscara para datos solo del baseline
 - 5: Aplicar máscara del baseline a PPMI
 - 6: Máscara para datos con MCI=1
 - 7: Máscara para datos con CU=1
 - 8: Máscara para datos con Dementia=1
 - 9: Introducción de PPMI_bl filtrados al MCI=1, PPMI_bl filtrados al CU=1 y PPMI_bl filtrados al Dementia=1
 - 10: **para** cada grupo clínico **hacer**
 - 11: **para** cada perfil **hacer**
 - 12: Filtrar los que poseen ese perfil
 - 13: Calcular porcentaje del grupo con ese perfil
 - 14: **fin para**
 - 15: Mostrar resultados de los perfiles del grupo al baseline
 - 16: **fin para**
 - 17: Representación de valores de $A\beta$ y pTau al baseline
 - 18: Representación de diagrama de barras de los perfiles según el estado cognitivo al baseline
 - 19: Introducción de PPMI filtrados al MCI=1, PPMI filtrados al CU=1 y PPMI filtrados al Dementia=1
 - 20: **para** cada grupo clínico **hacer**
 - 21: **para** cada perfil **hacer**
 - 22: Guardar valor mínimo del $A\beta$
 - 23: Guardar valor máximo del pTau
 - 24: Filtrar los que poseen ese perfil
 - 25: Calcular porcentaje del grupo con ese perfil
 - 26: **fin para**
 - 27: Mostrar resultados de los perfiles del grupo al longterm
 - 28: **fin para**
 - 29: Representación de valores de $A\beta$ y pTau al longterm
 - 30: Representación de diagrama de barras de los perfiles según el estado cognitivo al longterm
-

4.5. Análisis estadístico.

En el apéndice A se hace un desarrollo de conceptos importantes para el entendimiento de este y capítulos posteriores sobre conocimientos estadísticos.

Se procede en esta sección a realizar, en primer lugar, modelos predictivos. Estos modelos proporcionan vectores de características óptimos en la predicción de conversión a MCI. Se procede a la realización de un modelo de progresión de la enfermedad (DPM²) para establecer la trayectoria a lo largo del tiempo de la PD. Para ello, se sigue las indicaciones previas establecidas en Platero et al. [63].

²Disease Progression Model.

4.5.1. Realización de modelos predictivos.

Este modelo es un estudio longitudinal multivariante. Se tienen datos de distintos puntos a lo largo del tiempo (bl, m12, m24) con diversas características de MRI, evaluaciones cognitivas, imágenes DAT, marcadores CSF y test motóricos. Se realiza el análisis de supervivencia con el modelo de Cox y de Cox extendido.

Se realiza mediante el *script* `builtPredictiveModels_PDtoPDMCI_Magerit_v1.m` para la construcción de los modelos en el directorio `/demo/modelos_predicitvos_PPMI/`.

En este proceso se realizan dos bucles anidados para la validación cruzada de k-folds (Figura 4.1):

- **Bucle interno:** Selecciona los subconjuntos de características óptimas con el algoritmo mRMR. Primero realiza validación cruzada dividiendo los datos en *training set* (90 %) con análisis y *test set* (10 %) que lo valida. Después se ejecuta el mRMR sobre el training y se seleccionan las características que aparecen con mayor frecuencia para elaborar los modelos predictivos.
- **Bucle externo:** Se realiza la misma división. El grupo test evalúa los modelos del bucle interno.

Construidos los modelos, se obtienen archivos `.mat` para cada punto en el tiempo y los tipos de datos elegidos. Se guarda un archivo cada 50 iteraciones hasta 1000.

Para trabajar con los `.mat`, se emplea el *script* `analysis_withSavedData_Mux_PPMI.m`. Este algoritmo genera un archivo `better_... .dat`. Tras generarlo, se pueden mostrar los datos de sensibilidad, especificidad, área bajo la curva y frecuencia de los distintos vectores de características para poder seleccionar el mejor de ellos. Al igual que en el caso anterior, se encuentra en el directorio `/demo/modelos_predicitvos_PPMI/`.

`builtPredictiveModels_PDtoPDMCI_Magerit_v1.m`

Para configurar los modelos predictivos para el algoritmo 9, se requieren diferentes parámetros:

- **n_Approach:** Selección del método de elaboración del modelo predictivo.
 - Modelo de Cox extendido.
 - Modelo exclusivo LME.
 - Combinación de los dos anteriores.
- **n_Cov:** Número de covariables.
 - Edad. Elaboración de efectos fijos en el LME con solo la edad y el campo magnético de la MRI.
 - Edad, sexo y nivel educativo. Elaboran con estas covariables y el campo magnético de la MRI.
 - Ninguna. Por defecto, utiliza la edad y el campo magnético de la MRI.

- **n_typeFeat:** Tipo de características para entrenar los modelos:
 - Regiones de Interés (ROI³).
 - Resultados de test neuropsicológicos.
 - Combinación de ROI y test neuropsicológicos.
 - Combinación de ROI, test neuropsicológicos, medidas CSF y resultados PET.
- **n_Population:** Población de entrenamineto del modelo.

Algoritmo 9 builtPredictiveModels_PDtoPDMCI_Magerit_v1

Entrada: Tabla Tresults.

Salida: Vectores de características con mayor eficacia

```

1: Configuración de los modelos predictivos
2: Obtener Tresults
3: Establecer efectos fijos del lote de entrenamiento
4: si LME entonces
5:   si Análisis de supervivencia entonces
6:     Obtener matriz de características
7:   si no
8:     Obtener matriz de características
9:   fin si
10: si no
11:   No matriz de características
12: fin si
13: para i=1 hasta i=totales del bucle externo hacer
14:   para i=1 hasta i=totales del bucle interno hacer
15:     Dividir la población en los dos lotes buscando el conjunto más óptimo.
16:     Entrenar el modelo.
17:   fin para
18:   Estimar el rendimiento del modelo entrenado con el lote de entrenamiento
19: fin para
20: Guardar vectores de características de rendimiento óptimo

```

analysis_withSavedData_Mux_PPMI.m

Tras el algoritmo anterior, se seleccionan los vectores de características de mejor rendimiento, sensibilidad, especificidad, AUC y frecuencia (algoritmo 10).

Algoritmo 10 analysis_withSavedData_Mux_PPMI

Entrada: Modelos predictivos del algoritmo 9

Salida: Características con mejor rendimiento

```

1: Configuración de los modelos predictivos
2: si NO EXISTE fichero con mejores características entonces
3:   Obtener rendimientos de cada vector de características
4:   Guardar resultados en tabla de Matlab
5: si no
6:   Cargar el fichero
7:   Imprimir en pantalla los vectores de características óptimas y sus propiedades.
8: fin si

```

³Region Of Interest

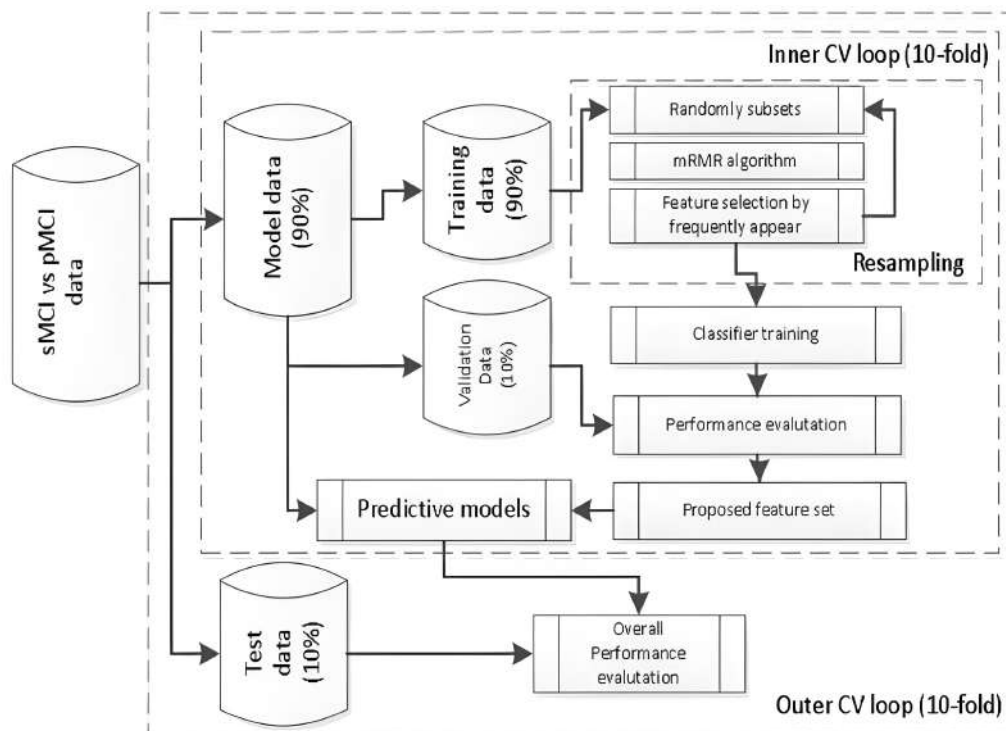


Figura 4.1: Procedimiento anidado de validación cruzada de 10 veces para el desarrollo y evaluación de modelos.

4.5.2. Realización de modelos de progresión. GRACE

Después de la obtención de características, se procede al modelado de la trayectoria temporal de la progresión de la patología. Se recurre al procedimiento descrito en Donohue et al. [34] para conseguir el conjunto de características que ofrece una descripción más definida de la evolución temporal del promedio de sujetos.

El método empleado es el método GRACE⁴, especializado en la estimación de curvas de crecimiento multivariantes en periodos prolongados de tiempo. Este método se encuentra el repositorio de Donohue et al. [34].

<https://bitbucket.org/mdonohue/grace/src/master/>

El modelo GRACE permite trazar trayectorias generales en el largo plazo⁵ a partir de trayectorias individuales a corto plazo⁶. El análisis de las trayectorias *short-term* emplea la medida de γ como el tiempo relativo a la evolución de la enfermedad.

$$Y_{ij} = g_j(t + \gamma_i) + \alpha_{0ij} + \alpha_{1ij}t + \epsilon_{ij}(t) \quad (4.1)$$

Donde:

- Y_{ij} : vector de resultados del sujeto i en el tiempo de medida j .

⁴GRowth models via Alternating Conditional Expectation

⁵Long-term

⁶short-term

- g_j : función diferenciable dependiente del tiempo de observación de estudio t y que incluye para cada sujeto un desfase temporal γ_i .
- α_{0ij} : vector de los efectos aleatorios independientes del tiempo.
- α_{1ij} : vector de los efectos aleatorios dependientes del tiempo.
- $\epsilon_{ij}(t)$: error residual.

Para aplicar la ecuación, se utiliza el algoritmo 11. Este algoritmo estima el valor de los parámetros g_i , γ_i y α a partir de otros parámetros por convergencia de residuos parciales. Los residuos parciales se estiman por el vector $(R_{ij}^g(t), R_{ij}^\gamma(t), R_{ij}^\alpha(t))$.

Algoritmo 11 GRACE

Entrada: Definición de cálculo de residuos parciales

Salida: Estimación de parámetros g_i , γ_i y α

- 1: Inicializar γ_i a cero
 - 2: Igualar parámetros α a cero
 - 3: **mientras** NO CONVERGAN $R_{ij}^g(t), R_{ij}^\alpha(t)$ **hacer**
 - 4: Estimar g_j suavizando el residuo parcial $R_{ij}^g(t)$
 - 5: Estimar α a partir del LME de $R_{ij}^\alpha(t)$
 - 6: Comprobar convergencia de $R_{ij}^g(t)$ y $R_{ij}^\alpha(t)$
 - 7: **fin mientras**
 - 8: **mientras** NO CONVERGAN $R_{ij}^g(t), R_{ij}^\alpha(t)$ y $R_{ij}^\gamma(t)$ **hacer**
 - 9: **mientras** NO CONVERGAN $R_{ij}^g(t), R_{ij}^\alpha(t)$ **hacer**
 - 10: Estimar g_j suavizando el residuo parcial $R_{ij}^g(t)$
 - 11: Estimar α a partir del LME de $R_{ij}^\alpha(t)$
 - 12: Comprobar convergencia de $R_{ij}^g(t)$ y $R_{ij}^\alpha(t)$
 - 13: **fin mientras**
 - 14: Tomar estimación de g_j
 - 15: Igualar a cero los parámetros α y el error $\epsilon_{ij}(t)$
 - 16: Estimar $R_{ij}^\gamma(t)$ para todo tiempo e instante
 - 17: Comprobar convergencia de los tres residuos
 - 18: **fin mientras**
 - 19: Asumir el valor de los parámetros g_i , γ_i y α equivalente a los resultados de las estimaciones de los residuos.
-

El cálculo de GRACE se realiza por tres ficheros. Se puede extraer el Modelo de Progresión de la Enfermedad (DPM^7). Se realiza en los siguientes pasos:

1. Generación de una tabla EXCEL mediante el *script tableGracePPMI_2.m* con los siguientes datos:
 - **Tiempo normalizado** (t): Momento de medida del marcador examinado.
 - **Y**: Vector de marcadores expresados como percentiles.
 - **raw**: Marcadores sin cambios.
 - **Outcome**: Número de identificación del marcador.
 - **id**: Número de identificación del sujeto.
 - **Convert**: Valor booleano que marca la conversión a MCI del sujeto.
 - **Convert/Censuring Time**: Tiempos de conversión y censura.

⁷Disease Progression Model

- **MCI:** Grupo clínico del sujeto (MCI o demencia).

En la creación de esta tabla ha realizado:

- Ajuste de datos a las restricciones estipuladas por GRACE.
 - Establecer la media de los tiempos de visita de cada sujeto a cero.
 - Corregir la edad del sujeto en cada visita.
 - Transformar el valor de cada marcador a una escala percentil.
2. Con la tabla EXCEL creada en el *script* anterior se invoca a RStudio para emplear el código de Donohue et al. [34]. Se utiliza el fichero *grace.PPMI.R*. Realiza el modelado de la enfermedad. Obtienen una nueva tabla con los parámetros de la ecuación anterior ya calculados.
 3. La tabla generada en el *script* mediante RStudio es la entrada del *script plot_gracePPMI_2.m*, que se encarga del trazado de las trayectorias a largo plazo y corto plazo de la población. Se calculan como principales parámetros:
 - Error Medio Absoluto (MAE⁸) de los residuos de estimación de trayectorias para cada marcador.
 - Rendimiento del modelo entrenado según grupo clínico.
 - Correlación entre el tiempo de conversión en GRACE y el conocido. Con esto se estima la validación del modelo predictivo.
 - Gráficos de las trayectorias.

Realización de modelos de progresión de forma masiva con Magerit

Se incorpora un nuevo modo de encontrar cuál es el vector de características que nos aporta un mejor rendimiento en la predicción del progreso de la enfermedad de Parkinson.

Para ello, se basa en una combinación de vectores de características y en la prueba de todas estas combinaciones posibles para clasificarlas y así obtener el vector más óptimo de todos.

Este método no se basa en estudios anteriores, en modelos predictivos generados o en análisis al inicio del estudio que ofrezca variables con importantes discrepancias.

Para obtener todas las combinaciones posibles, ejecutar GRACE en todas esas combinaciones y obtener la clasificación con los mejores resultados se genera un *script* para su ejecución en Magerit y reducir el tiempo de ejecución y la obtención de resultados.

El código fuente base para la realización de este método ha sido establecido por Rodrigo Batal, aunque este ha sufrido modificaciones para su uso con PPMI y su optimización.

⁸Medium Absolute Error

Este algoritmo (12) permite ejecutar los tres pasos en solo un código. Además, ahorra tiempo comprobando si el trabajo que se va a realizar ya ha sido ejecutado previamente y permitiendo seleccionar cual de los tres pasos se quiere realizar según las necesidades o preferencias del usuario.

El *script* se encuentra en */demo/multi_grace_PPMI/multi_grace_PPMI.m*

Algoritmo 12 multi_grace_PPMI

Entrada:**Salida:** Clasificación de los vectores de características con mejor rendimiento

- 1: Establecer población, tipos de características y criterio MCI
 - 2: Establecer o no la edad como covariable
 - 3: Crear directorio para resultados
 - 4: Selección de características a combinar
 - 5: Selección de longitud de vectores de características
 - 6: Selección de pasos a realizar
 - 7: Extracción de población PD y criterio MCI
 - 8: Establecer año cero
 - 9: Introducir la covariable de la edad o no
 - 10: Realización de las diferentes combinaciones de vectores
 - 11: Paso a percentiles
 - 12: Comprobar array de vectores y si no se produce su creación
 - 13: **si** phase(1)=TRUE **entonces**
 - 14: Creación de tablas *dd... .xlsx*.
 - 15: **fin si**
 - 16: **si** phase(2)=TRUE **entonces**
 - 17: Realización de algoritmo 11 y generación de tabla *dd1... .xlsx*
 - 18: **fin si**
 - 19: **si** phase(3)=TRUE **entonces**
 - 20: Extracción de tablas *dd1... .xlsx*
 - 21: Generación de resultados *longterm*
 - 22: **fin si**
 - 23: Muestreo de los mejores vectores de características de mayor a menor rendimiento
 - 24: Creación de tabla *.xlsx* para guardar los resultados obtenidos.
-

Capítulo 5

Resultados

5.1. Introducción

En este capítulo se presentan los resultados obtenidos en la realización del trabajo con su justificación e importancia en el estudio realizado.

En primer lugar, se comienza con el análisis de los resultados obtenidos al inicio del estudio, así como los tiempos de conversión y censura obtenidos en los diferentes grupos clínicos y estados cognitivos de la PD. Además se le suma el análisis de los perfiles AT de la AD y los factores de riesgo y predictores de la enfermedad.

En segundo lugar, tras este análisis transversal, se procede al análisis de los resultados de forma longitudinal con la elaboración de los modelos predictivos. Se empieza con el análisis de modelos univariantes, continuando con multivariantes de solo características neuropsicológicas y modelos con todas las características que se disponen.

En tercer lugar, se presentan los resultados de los modelos de progresión con los vectores de características obtenidos en los modelos predictivos y a través del estudio e investigación de características que discrepan al inicio de estudio, además se comparan con el método masivo de combinación de características ejecutado en Magerit, buscando el mejor rendimiento.

5.2. Resultados del análisis al inicio del estudio

5.2.1. Análisis al inicio del estudio de sujetos PD y HC

Tal y como se observa en la tabla 5.1, la mayoría de los valores de los resultados cognitivos tienen diferencias significativas ($p \leq 0.001$), exceptuando LNS, JLO y ADL ($p=0.3381$, $p=0.0495$ y $p=0.2118$, respectivamente). Por otro lado, los marcadores de $A\beta$, aSyn, Tau y pTau son más bajos en los sujetos PD ($n=423$) que en los HC ($n=196$) (figura 5.1) y con diferencias significativas ($0.0001 < p < 0.0209$). Por otra parte, no hay diferencias significativas en los convertidores APOE4 entre sujetos PD y HC (26.3 % y 26,4 %, respectivamente) (tabla 5.1).

Tabla 5.1: Características demográficas y clínicas basales de los pacientes con HC y PD a los 5 años de seguimiento. Los datos son la media (desviación estándar) (mínimo, máximo) o n (%).

Características demográficas y clínicas		
Sujetos	423 (PD)	196 (HC)
Mujeres	146 (34.5%)	70 (35.7%)
APOE4	101 (26.3%)	47 (26.4%)
Edad	61.66 (9.71)	60.82 (11.23)
Educación	15.56 (2.97) (5.00 26.00)	16.04 (2.89) (8.00 24.00)
GDS	2.32 (2.44) (0.00 14.00)	1.29 (2.10) (0.00 15.00)
UPSIT	22.35 (8.23) (1.00 40.00)	33.98 (4.86) (11.00 40.00)
REM	4.12 (2.69) (0.00 12.00)	2.83 (2.26) (0.00 11.00)
UPDRS3	20.89 (8.85) (4.00 51.00)	1.21 (2.19) (0.00 13.00)
Resultados cognitivos		
MOCA	27.13 (2.32) (17.00 30.00)	28.23 (1.11) (26.00 30.00)
LNS	10.59 (2.66) (2.00 20.00)	10.87 (2.57) (2.00 20.00)
SDMT	41.18 (9.73) (7.00 82.00)	46.77 (10.53) (20.00 83.00)
JLO	12.77 (2.13) (5.00 15.00)	13.12 (1.98) (4.00 15.00)
INMEDIAT	24.44 (4.98) (9.00 36.00)	26.05 (4.50) (15.00 35.00)
HVLTRDLY	8.36 (2.52) (0.00 12.00)	9.29 (2.32) (2.00 12.00)
SFT	48.67 (11.63) (20.00 103.00)	51.80 (11.20) (22.00 80.00)
ADL	93.14 (5.89) (70.00 100.00)	100.00 (0.00) (100.00 100.00)
SCOPA	9.51 (6.18) (0.00 39.00)	5.83 (3.69) (0.00 20.00)
Marcadores de CSF e imágenes de DAT		
ABETA	909.55 (410.73) (238.80 3707.00)	1019.37 (499.46) (239.10 3297.00)
ASYN	1506.70 (666.59) (432.40 5256.90)	1695.19 (747.42) (488.60 4683.10)
TAU	169.54 (56.80) (80.93 467.00)	191.64 (79.26) (81.96 580.80)
PTAU	14.87 (5.25) (8.01 40.13)	17.52 (8.35) (8.08 73.61)
Abeta/TAU	5.56 (1.60) (1.19 10.38)	5.58 (1.67) (0.71 9.80)
MEANPUT	0.83 (0.30) (0.24 2.17)	2.15 (0.56) (0.64 3.89)
MEANCAU	2.00 (0.56) (0.39 3.71)	2.98 (0.63) (1.32 5.20)

En cuanto a la correlación, entre $aSyn$ y $A\beta$ no es tan alta tanto en PD como HC ($R=0.63$). En la relación $aSyn$ y $pTau$ si que hay una mayor correlación en ambos grupos ($R=0.78-0.85$). Destaca la alta correlación entre los indicadores $pTau$ y Tau ($R=0.95-0.96$) (figura 5.2).

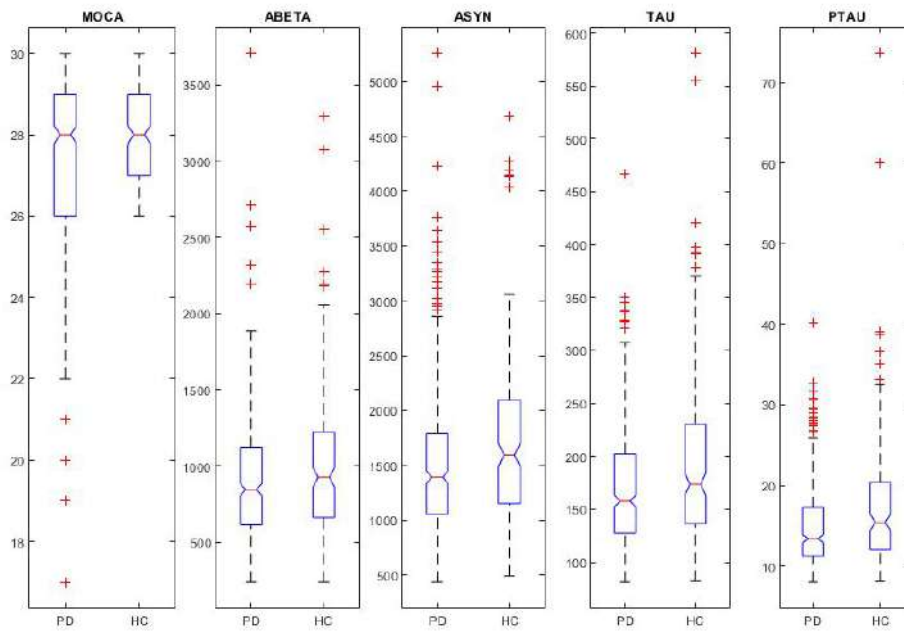


Figura 5.1: Valores al inicio del estudio entre sujetos PD y HC.

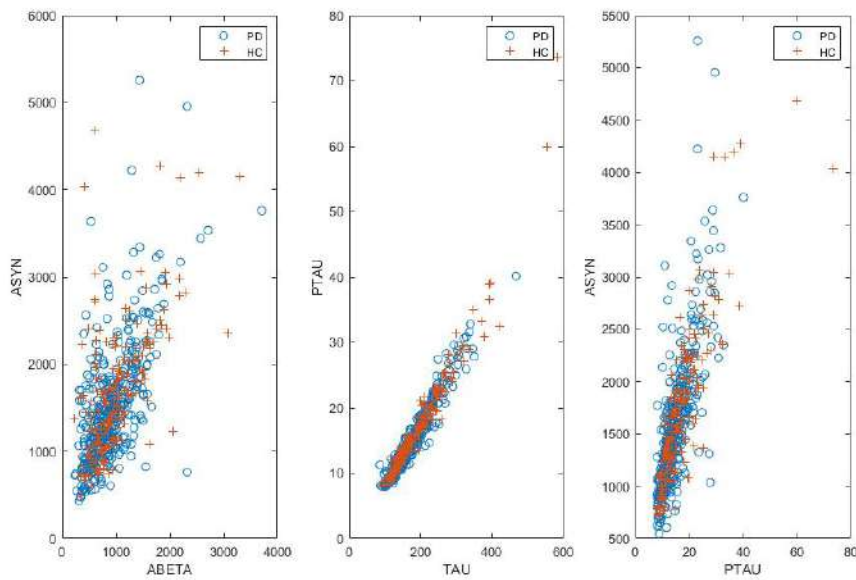


Figura 5.2: Correlación entre marcadores de CSF y aSyn al inicio del estudio entre PD y HC.

5.2.2. Análisis al inicio del estudio de sujetos PD-A+, PD-A-, HC-A+ y HC-A-

Características demográficas y clínicas				
Sujetos	281 (PD-A-)	129 (PD-A+)	136 (HC-A-)	52 (HC-A+)
Mujeres	99 (35.2%)	41 (31.8%)	50 (36.8%)	18 (34.6%)
APOE4	53 (20.7%)	47 (40.2%)	22 (17.5%)	23 (51.1%)
Edad	61.47 (9.63)	62.22 (9.64)	60.67 (10.79)	61.00 (12.98)
Educación	15.58 (2.93)	15.50 (3.06)	16.09 (2.99)	15.87 (2.78)
Resultados cognitivos				
MOCA	27.06 (2.33)	27.33 (2.23)	28.21 (1.10)	28.27 (1.14)
LNS	10.51 (2.74)	10.68 (2.44)	10.91 (2.71)	10.98 (2.24)
HVLTRDLY	8.36 (2.46)	8.29 (2.68)	9.35 (2.46)	8.98 (2.03)
Marcadores de CSF e imágenes de DAT				
ABETA	1089.55 (371.11)	517.47 (108.31)	1202.85 (466.72)	539.51 (116.43)
ASYN	1673.37 (673.47)	1135.95 (492.12)	1798.87 (682.25)	1413.84 (851.97)
TAU	180.19 (53.12)	143.16 (58.04)	198.07 (64.84)	173.39 (109.82)
PTAU	15.42 (4.95)	13.21 (5.86)	17.27 (6.10)	18.44 (13.72)

Tabla 5.2: Características demográficas y clínicas basales de los pacientes con HC-A+ y HC-A- y PD-A+ y PD-A- a los 5 años de seguimiento. Los datos son la media (desviación estándar) o n (%).

Si se refiere a la población amiloide (tabla 5.2), en ambos grupos es similar con un 31.46 % (n=129/410) en sujetos PD y un 27.66 % (n=129/188) en sujetos HC. El criterio para la clasificación como A+ ha sido de $A\beta_{1-42} < 683.45$. Se mantienen una total diferencia significativa en $A\beta$, aSyn, Tau y pTau (p=0.000) con valores más altos en PD-A- que en PD-A+ (figura 5.3). En cambio, en los sujetos HC, aunque hay mayor diferencia significativa en $A\beta$, aSyn y Tau (p<0.001), no es tanto en pTau (p=0.1203) (Figura 5.4). Si se fija atención en MoCA, no existen discrepancias entre PD-A- y PD-A+ ni entre HC-A- y HC-A+. También, fijándose en APOE4, es mayor en los sujetos A+ que en los A- tanto en PD como en HC. En cuanto a los resultados cognitivos, no hay diferencias significativas entre A+ y A-.

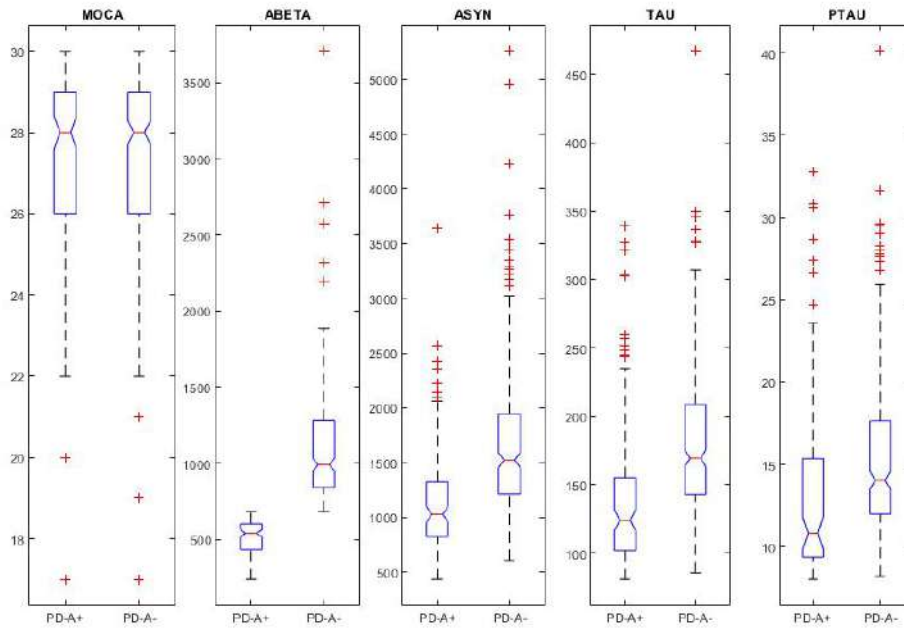


Figura 5.3: Valores al inicio del estudio entre sujetos PD-A+ y PD-A-.

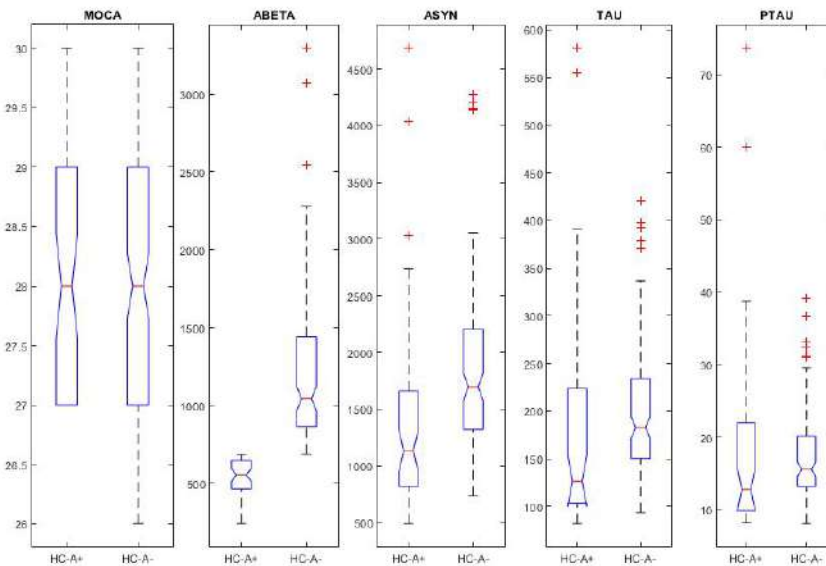


Figura 5.4: Valores al inicio del estudio entre sujetos HC-A+ y HC-A-.

5.2.3. Análisis al inicio del estudio de sujetos PD-CU y PD-MCI y PDD.

Características demográficas y clínicas				
Sujetos	323 (PD-CU)	83 (PD-MCI)	17 (PD-Dementia)	Multicompare
Mujeres	120 (37.2%)	20 (24.1%)	6 (35.3%)	
APOE4	78 (26.4%)	18 (23.7%)	5 (38.5%)	
Edad	60.13 (9.82) (33.50,82.259)	66.01 (7.37) (48.03,84.88)	69.45 (7.45) (54.82,82.98)	
Educación	15.70 (2.87) (8.00 26.00)	14.99 (2.82) (8.00 21.00)	15.59 (4.91) (5.00 21.00)	-
GDS	2.17 (2.40) (0.00 14.00)	2.86 (2.61) (0.00 13.00)	2.71 (2.11) (0.00 8.00)	-
UPSIT	23.41 (7.87) (5.00 39.00)	19.54 (8.60) (1.00 40.00)	16.06 (7.39) (5.00 31.00)	ac
REM	3.82 (2.54) (0.00 12.00)	5.06 (2.90) (0.00 12.00)	5.24 (3.11) (0.00 12.00)	a
UPDRS3	20.51 (8.73) (5.00 47.00)	22.29 (9.40) (4.00 51.00)	21.29 (8.18) (10.00 39.00)	-
Resultados cognitivos				
MOCA	27.33 (2.17) (19.00 30.00)	26.58 (2.70) (17.00 30.00)	25.94 (2.38) (21.00 29.00)	ac
LNS	11.04 (2.50) (2.00 20.00)	9.29 (2.58) (4.00 17.00)	8.29 (2.78) (3.00 13.00)	ac
SDMT	42.94 (8.75) (17.00 82.00)	36.69 (10.18) (7.00 58.00)	29.82 (10.85) (15.00 50.00)	abc
JLO	12.98 (1.96) (5.00 15.00)	12.19 (2.44) (6.00 15.00)	11.53 (2.67) (6.00 15.00)	ac
INMEDIAT	25.26 (4.74) (9.00 36.00)	22.13 (4.90) (11.00 34.00)	20.24 (4.28) (14.00 28.00)	ac
HVLTRDLY	8.75 (2.35) (0.00 12.00)	7.13 (2.75) (0.00 12.00)	7.00 (2.12) (3.00 11.00)	ac
SFT	50.60 (11.54) (24.00 103.00)	43.48 (9.39) (20.00 70.00)	37.35 (9.37) (25.00 54.00)	ac
ADL	93.25 (6.02) (70.00 100.00)	93.01 (5.29) (80.00 100.00)	91.76 (6.11) (80.00 100.00)	-
SCOPA	8.74 (5.70) (0.00 35.00)	11.79 (6.96) (3.00 39.00)	13.06 (7.20) (4.00 32.00)	ac
Marcadores de CSF e imágenes de DAT				
ABETA	942.40 (413.12) (315.90 3707.00)	844.08 (395.26) (238.80 2572.00)	592.27 (251.10) (308.80 1300.00)	c
ASYN	1519.17 (680.58) (432.40 5256.90)	1470.97 (620.62) (472.00 3443.70)	1443.49 (642.42) (539.40 2780.50)	-
TAU	167.17 (54.61) (80.93 467.00)	172.53 (59.27) (83.93 345.30)	202.68 (78.07) (97.17 327.40)	c
PTAU	14.52 (4.99) (8.02 40.13)	15.51 (5.40) (8.36 32.80)	18.59 (7.68) (8.01 30.84)	c
Abeta/TAU	5.78 (1.52) (1.38 10.38)	5.14 (1.54) (1.19 8.08)	3.24 (1.38) (1.39 6.30)	abc
MEANPUT	0.85 (0.29) (0.24 2.15)	0.77 (0.32) (0.29 2.17)	0.66 (0.18) (0.31 1.05)	c
MEANCAU	2.07 (0.54) (0.39 3.71)	1.80 (0.54) (0.62 3.27)	1.59 (0.50) (0.74 2.52)	ac

Tabla 5.3: Características demográficas y clínicas basales de los pacientes con PD-CU, PD-MCI y PDD a los 5 años de seguimiento con el criterio MCIttest. Los datos son la media (desviación estándar) (mínimo, máximo) o n (%).

En el análisis de los diferentes grupos usando el criterio SITE (tabla 5.3), se observa que los hombres son los que más padecen la PD (64.7%), pero sin unas diferencias altamente significativas. Se puede concluir además, que la edad es un factor importante en el desarrollo de la enfermedad aumentando desde una media de 60.1 años en los sujetos PD-MCU a 66.0 años de edad en PD-MCI y hasta 69.4 años de edad en PDD.

En MoCA no se observan discrepancias entre los tres grupos (27.33, 26.58 y 25.94 en los sujetos PD-CU, PD-MCI y PDD, respectivamente). Si se refiere a los marcadores CSF y marcadores de imagen DAT, se puede dar importancia la $A\beta$ que disminuye desde PD-CU (942.40) a los sujetos PDD (592.270) o a Tau y pTau que aumenta desde PD-CU a PDD.

Se ha utilizado finalmente el análisis con `anova1` y la multicomparación para enfatizar en que características existen discrepancias y establecer como factores de riesgo. El significado de las letras es el siguiente:

- a: Discrepancia entre PD-CU y PD-MCI.
- b: Discrepancia entre PD-MCI y PDD.
- c: Discrepancia entre PD-CU y PDD.

Con este significado establecido, se puede observar como la medida cognitiva SDMT presenta discrepancias entre los tres grupos cognitivos (PD-CU, PD-MCI y

PDD) de los sujetos que poseen la PD con valores de 42.96, 36.36 y 29.82, respectivamente. Si se fija atención además, en los marcadores de imágenes DAT y CSF, se pueden observar discrepancias también en los tres grupos cognitivos en la medida de $A\beta$ /Tau con valores de 5.78, 5.14 y 3.24, respectivamente.

La mayoría de medidas cognitivas, excepto ADL, presenta discrepancias entre PD-CU con PD-MCI o PDD. A esto, se le añade el mismo tipo de discrepancia a la medida del caudado (MEANCAU) con valores de 2.07, 1.80 y 1.59, respectivamente.

De este modo, se pueden establecer como factores de riesgo más influyentes y que discrepan en el progreso del estado cognitivo en la enfermedad de Parkinson a la medida SDMT y $A\beta$ /Tau junto con la edad.

Por ello, son medidas que nos encontraremos en vectores de modelos predictivos de la enfermedad y en su análisis en los modelos de progresión.

Tiempos de conversión y censura

En la figura 5.5 se puede observar que la población obtenida de PPMI a los 5 años, $n=83/423$ (20 %) convierten a MCI y que $n=17/423$ (4 %) lo hacen a demencia. De la población PD, $n=25/423$ (6 %) pasa a formar parte de PD-MCI y $n=5/423$ (1 %) lo hace a demencia en el primer año.

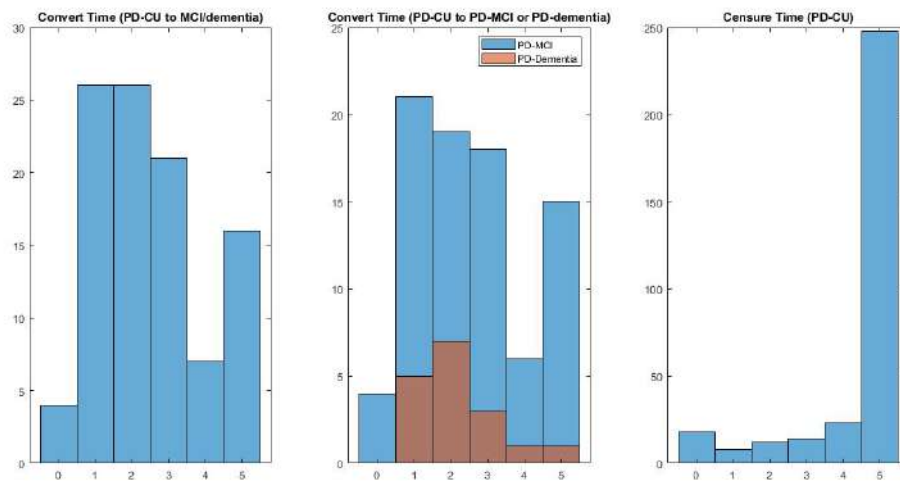


Figura 5.5: Histograma de la población PD atendiendo al tiempo de conversión y censura. En la de la izquierda, conversión a MCI o demencia. En la central, separación de conversión entre MCI y demencia. En la de la derecha, sujetos no convertidos.

5.2.4. Predictores y factores de riesgo

En la tabla 5.4 se observan otros distintos factores de riesgo (APOE4, A+, género y REM) y predictores de conversión ($MoCA < 24$, $PACC < -1.0$) al inicio del estudio

que se han establecido al analizar los diferentes grupos clínicos sin análisis de multicomparación y en los grupos cognitivos de los sujetos con PD según los diferentes estudios y artículos analizados anteriormente.

Se puede observar que el APOE4 aumenta con la progresión de la enfermedad a demencia (23.4 % y 26.7 % en PD-CU y PDD, respectivamente). En el factor de riesgo que si se observan diferencias significativas, es la patología amiloide (A+, $A\beta < 683.45\text{pg/ml}$) con una relación de 23.4 % para PD-CU y un 66.7 % para PDD.

También es observable en el predictor MoCA < 24 siendo el doble de sujetos en proporción entre PD-CU y PDD (6.5 % y 13.3 %), aunque sin aparentes diferencias entre PD-MCI y PDD (14.7 % y 13.3 %). En el predictor que si se observa una discrepancia relevante es el indicador PACC < -1.0, observando un aumento del 22.91 % en PD-CU hasta un 70.59 % en PDD.

Clinical	General population with CSF markerscat baseline						
	n	MOCA < 24	APOE4	A+	PACC < -1.0	Female	REM
PDD	17	13.3 %	26.7 %	66.7 %	70.59 %	40.0 %	46.7 %
PD-MCI	83	14.7 %	23.5 %	26.5 %	51.81 %	26.5 %	54.4 %
PD-CU	323	6.5 %	23.4 %	23.4 %	22.91 %	36.1 %	32.6 %

Tabla 5.4: Resultados de conversión a MCI o demencia respecto a los seis altos factores de riesgo (MOCA < 26, APOE4, A+, PACC, Female y REM) al inicio del estudio mediante clasificación de CU, MCI y demencia.

5.2.5. Perfil AT en sujetos PD

Se establece, como se ha referido anteriormente, a la patología amiloide (A+) como $A\beta < 683.45\text{pg/ml}$. Se refiere a T+ como $p\text{Tau} > 23\text{pg/ml}$.

Al inicio del estudio, el perfil A+ se establece de la siguiente forma:

- 23.4 % en PD-CU.
- 26.5 % en PD-MCI.
- 66.7 % en PDD.

Se puede observar que 2/3 de la población PDD (1 % de la población PDD en el primer año) presenta un perfil A+ al inicio del estudio, antes de convertir. Ese 66.7 % es mayor al 50 % de los sujetos con PDD que se encuentran en Alzheimer continuo. Tal y como se ha comentado previamente, no hay diferencias significativas entre PD-CU y PD-MCI. Estos resultados coinciden con la estimación de que un 30 % de los sujetos PD-CU y PD-MCI son A+ al inicio.

En el análisis longitudinal se establece el perfil A+ con el mínimo valor en los 5 años, dando como resultado los siguientes perfiles A+:

- 36.7 % en PD-CU.
- 41.9 % en PD-MCI.

- 66.7% en PDD.

Si se establece el análisis en la patología tau junto con la patología amiloide, se pueden realizar diversos tipos de análisis. Hay un 70.8% de los sujetos PD-CU que no tiene ni patología amiloide, ni tau (A-T-), siendo un 67.6% en el caso de PD-MCI y disminuyendo hasta un 13.3% en el caso de PDD.

Un 67.7% de los sujetos PDD están en Alzheimer continuo (A+) al inicio sin diferencias entre el análisis longitudinal y transversal. Los sujetos PD-MCI comienzan con 26.5% en A+, llegando a un 41.9% de forma longitudinal. En cambio, en los sujetos PD-CU, comienzan con A+ un 23.4% al inicio hasta un 36.7% en el estudio longitudinal. Estos resultados también se sitúan en la línea de que 1/3 de los PDD son A+T+ (30% aproximadamente), y más de la mitad están en A+ (66.7%).

En el caso de A-T+, es una mínima parte de los grupos, estableciéndose desde un mínimo de 5.8% en PD-CU al inicio del estudio hasta un 8.8% en PD-MCI en el análisis longitudinal para sujetos PD-CU y PD-MCI.

Clinical	time	A-T-	A-T+	A+	A+T+
PDD (n=17)	Baseline	13.3%	20.0%	66.7%	26.7%
	Longitudinal	13.3%	20.0%	66.7%	26.7%
PD-MCI (n=83)	Baseline	67.6%	5.9%	26.5%	2.9%
	Longitudinal	50.0%	8.1%	41.9%	8.1%
PD-CU (n=323)	Baseline	70.8%	5.8%	23.4%	0.7%
	Longitudinal	54.3%	9.0%	36.7%	2.0%

Tabla 5.5: Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores al inicio del estudio y en el análisis longitudinal.

Todos estos resultados pueden ser vistos de forma gráfica en las figuras 6.3 al inicio del estudios y 6.4 en el análisis longitudinal.

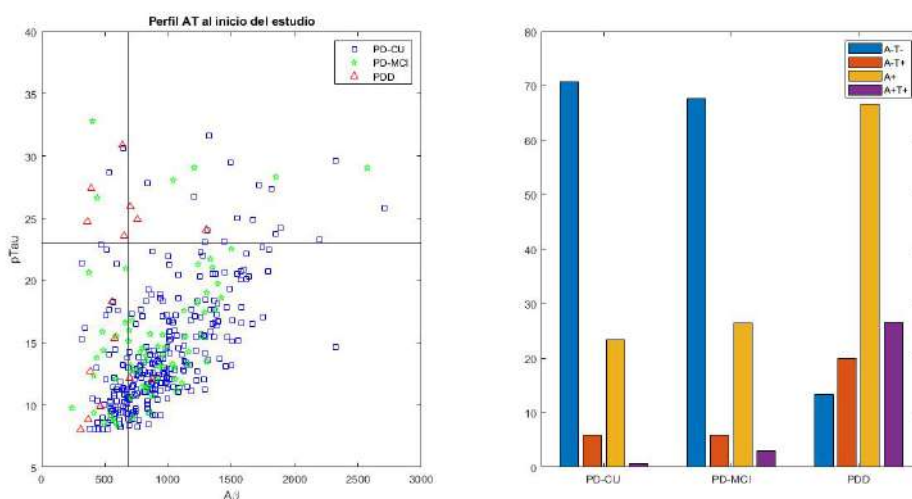


Figura 5.6: Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\tau_{ou} > 23$ pg/ml. Corresponde con los valores al inicio del estudio. Las líneas muestran los puntos de umbral en las medidas.

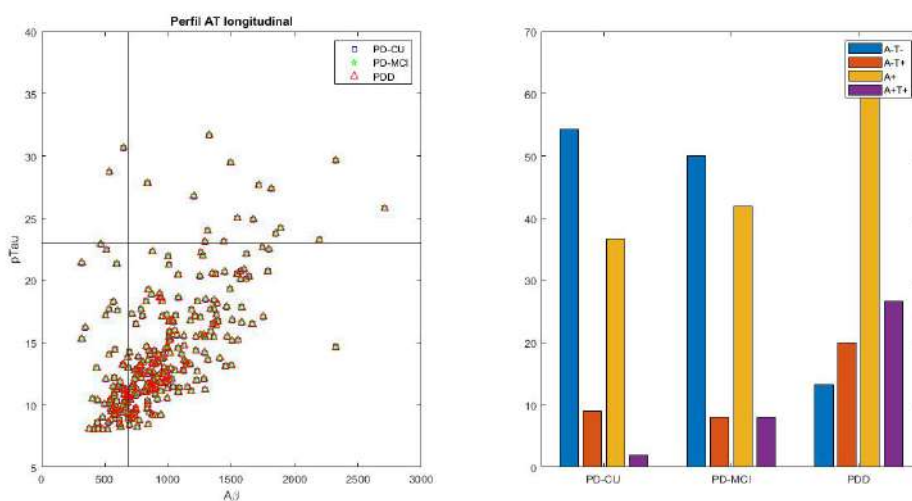


Figura 5.7: Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\tau_{ou} > 23$ pg/ml. Corresponde con los valores en el análisis longitudinal. Las líneas muestran los puntos de umbral en las medidas.

5.3. Resultados del análisis multivariante.

Se presentan los resultados obtenidos con el análisis multivariante para la realización de modelos predictivos que nos indican los biomarcadores con una mejor capacidad discriminadora entre los sujetos PD-MCI- y PD-MCI+.

5.3.1. Modelos predictivos.

Se utiliza el análisis de supervivencia. Se realizan medidas en las visitas *baseline*, mes 12 (primer año) y mes 24 (segundo año).

En las tablas se observan la selección de marcadores óptimos con las siguientes características:

- **Modelo empleado:** Tipo de análisis (*Cox* o *extCox*), criterio escogido (*cogstate* o *MCI testcores*) y vista examinada (*bl*, *m12*, *m24*).
- **Sensibilidad (SEN¹):** Verdaderos positivos clasificados, es decir PD-MCI+ detectados.
- **Especificidad (SPE²):** Verdaderos negativos clasificados, es decir, PD-MCI- detectados.
- **Exactitud (ACC³):** Exactitud de la clasificación.
- **Característica Operativa del Receptor (ROC⁴):** Sensibilidad frente a la especificidad en un gráfico. El ROC ideal es en el que SPE y SEN son del 100 %.
- **Frecuencia:** Número de veces que aparece un vector de características.
- **Vector de características:** Conjuntos de características del análisis de los modelos predictivos.

La ponderación de los mejores vectores de características se realiza tomando en cuenta el equilibrio SEN-SPE. Tienen que ser lo más altos y parejos, junto con un elevada frecuencia de aparición. Se compara en el criterio de SITE para el diagnóstico MCI según el modelo Cox al inicio del estudio y Cox extendido en el primer y segundo año.

a. Modelos predictivos univariantes al inicio del estudio

En primer lugar, se van a analizar los resultados obtenidos en la generación de modelos predictivos de forma univariante, es decir, solo dependiendo de una sola característica.

Para ello, se aplica el análisis Cox al inicio del estudio y se añade la edad como covariable. En la tabla 5.6, se pueden observar los resultados obtenidos en cada una de las distintas variables.

Se pueden observar resultados de las medidas REM, LNS, SDMT, INMEDIANTE, SFT y MEANCAU con medidas de sensibilidad, especificidad y precisión importantes (ACC=62.4%, 63.7%, 63.0%, 63.9%, 63.4% y 61.6%, respectivamente). También presentan una ROC adecuada al tratarse únicamente de una variable

¹Sensibility.

²Specifity.

³Accuracy.

⁴Receiver Operating Characteristic

Cox, bl					
SEN	SPE	ACC	ROC	Ponderación	
74.4(71.5 77.2)	53.1(51.4 54.9)	58.2(56.7 59.7)	69.6(67.8 71.4)	0.20	MOCA
71.1(68.5 73.8)	56.2(54.5 58.0)	59.6(58.2 61.1)	70.9(69.4 72.4)	0.28	GDS
74.8(71.7 77.9)	56.7(54.9 58.5)	61.2(59.7 62.7)	71.3(69.5 73.2)	0.31	UPSIT
74.4(71.7 77.2)	58.8(57.0 60.7)	62.4(60.9 63.9)	74.1(72.3 75.8)	0.44	REM
72.3(69.5 75.2)	60.9(59.3 62.5)	63.7(62.2 65.1)	73.2(71.2 75.1)	0.54	LNS
68.4(65.4 71.5)	61.4(59.7 63.1)	63.0(61.6 64.4)	72.7(70.9 74.5)	0.65	SDMT
68.8(66.2 71.4)	57.2(55.5 59.0)	60.0(58.5 61.4)	70.2(68.6 71.9)	0.34	JLO
69.1(65.8 72.4)	62.1(60.3 63.8)	63.9(62.2 65.6)	73.0(71.0 75.0)	0.50	INMEDIATE
68.4(65.6 71.2)	58.9(57.3 60.6)	61.2(59.9 62.6)	72.1(70.4 73.8)	0.38	HVLTRDLY
74.8(72.0 77.6)	59.8(58.3 61.3)	63.4(62.0 64.8)	74.3(72.6 76.0)	0.65	SFT
75.9(73.1 78.8)	53.8(52.2 55.5)	59.2(57.8 60.7)	68.4(66.7 70.2)	0.08	ADL
72.9(70.1 75.8)	55.8(54.1 57.6)	60.0(58.5 61.4)	71.1(69.4 72.8)	0.31	SCOPA
72.9(70.1 75.8)	55.1(53.1 57.2)	59.5(57.7 61.2)	69.4(67.3 71.5)	0.25	MEANPUT
76.1(73.5 78.7)	57.2(55.5 58.8)	61.6(60.2 63.0)	72.4(70.7 74.1)	0.46	MEANCAU
75.2(72.2 78.2)	55.9(54.2 57.7)	60.5(59.1 62.0)	71.3(69.7 73.0)	0.38	ABETA
77.5(74.9 80.1)	53.9(52.2 55.6)	59.5(58.1 61.0)	69.8(68.0 71.5)	0.14	ASYN
75.1(72.3 77.8)	53.7(51.8 55.6)	58.9(57.4 60.5)	68.4(66.8 70.1)	0.05	TAU
75.3(72.3 78.2)	54.1(52.2 56.0)	59.1(57.6 60.6)	69.2(67.6 70.8)	0.09	PTAU
71.3(68.1 74.6)	59.3(57.6 61.0)	62.0(60.5 63.5)	72.2(70.5 73.9)	0.42	ABetaTAU
72.9(70.2 75.7)	54.8(53.0 56.6)	59.3(57.9 60.8)	69.1(67.4 70.8)	0.24	UPDRSTOT
73.6(70.7 76.4)	54.9(53.2 56.6)	59.4(58.0 60.8)	69.9(68.1 71.7)	0.24	UPDRSTOTON
73.3(70.1 76.5)	56.7(55.0 58.4)	60.8(59.5 62.2)	71.4(69.7 73.1)	0.31	UPDRS1
74.4(71.7 77.1)	55.5(53.8 57.3)	59.8(58.4 61.3)	69.7(68.0 71.5)	0.21	UPDRS2
75.4(72.1 78.8)	52.7(51.0 54.4)	58.3(56.9 59.8)	68.1(66.3 69.9)	0.09	UPDRS3
76.4(73.8 79.0)	52.5(50.8 54.2)	58.3(56.8 59.7)	68.0(66.2 69.7)	0.09	UPDRS3ON

Tabla 5.6: Resultado del análisis de modelos predictivos univariantes al inicio del estudio con el análisis Cox y la edad como covariable.

(ROC=74.1 %, 73.2 %, 72.7 %, 73.0 %, 74.3 % y 72.4 %, respectivamente).

Si además, se enfatiza en la aportación de estas medidas a la predicción del estado cognitivo, las medidas presentan un elevado porcentaje, siendo 44 %, 54 %, 65 %, 50 %, 65 % y 46 %, respectivamente.

Todos ellos cercanos e incluso superiores al 50 %. Los medidores más destacables son SDMT y SFT con una aportación del 65 % en ambos casos. El resto de medidas se encuentran por debajo de un 40 %. Con ello se destacan SDMT y SFT como principales medidas predictivas en los modelos univariantes.

b. Modelos predictivos multivariantes solo con medidas neuropsicológicas al inicio del estudio

A continuación, se procede a la realización de modelos predictivos multivariantes para analizar y comprobar como mejora la predicción de la enfermedad con un vector de características en lugar de un modelo univariante.

Para ello, en este caso, se van a utilizar solo medidas neuropsicológicas al inicio del estudio con análisis Cox y la edad como covariable (tabla 5.7).

Se observan resultados mejores respecto al caso de los modelos univariantes con una mejora en valores como la precisión (ACC=69.6 %, 70.1 %, 69.4 %, 70.3 % y 69.8 %). A ello, se le suma el aumento de los valores de ROC próximos al 80 % (ROC=78.7 %, 78.2 %, 78.1 %, 78.0 % y 77.7 %).

Si además observamos la frecuencia de aparición de los vectores de características, estas superan el número 800, teniendo un máximo de 2640 el vector de características formado por LNS, SDMT, JLO, SFT y SCOPA junto con la edad como covariable.

Con esto, se demuestra como el desarrollo de modelos predictivos multivariantes, aunque solo con medidas neuropsicológicas mejora considerablemente a los modelos predictivos univariantes generados anteriormente.

Cox, Ag, bl						
SEN	SPE	ACC	ROC	Frecuency		
73.4(72.4 74.4)	68.4(67.8 69.0)	69.6(69.1 70.1)	78.7(78.2 79.2)	810	LNS,SDMT,JLO,HVLTRDLY,SCOPA	
73.0(72.4 73.7)	69.1(68.7 69.5)	70.1(69.7 70.4)	78.2(77.8 78.5)	2022	LNS,SDMT,JLO,INMEDIATE,SCOPA	
72.8(72.2 73.4)	68.3(68.0 68.7)	69.4(69.1 69.7)	78.1(77.8 78.5)	2260	LNS,SDMT,HVLTRDLY,SCOPA	
73.2(72.5 73.9)	69.3(68.9 69.7)	70.3(69.9 70.6)	78.0(77.6 78.4)	1506	MOCA,LNS,SDMT,SFT,SCOPA	
73.0(72.5 73.6)	68.8(68.4 69.1)	69.8(69.5 70.0)	77.7(77.4 78.1)	2640	LNS,SDMT,JLO,SFT,SCOPA	

Tabla 5.7: Resultado del análisis de modelos predictivos multivariantes solo con medidas neuropsicológicas al inicio del estudio con el análisis Cox y la edad como covariable.

c. Modelos predictivos univariantes al año del inicio del estudio

En el primer año (mes 12), los resultados tienen un peor diagnóstico. Se puede observar una sensibilidad entre el 68.0 % y 69.3 % y una especificidad entre el 70.2 % y 71.5 % y una frecuencia máxima de 1154. En este caso, ya se ha aplicado el Cox extendido y todos los vectores se encuentran entre 7 y 10 elementos. Se establece el vector de características que se comparte en todos los vectores escogidos y que son:

LNS, SDMT, SFT, MEANCAU, ABeta/TAU, UPDRS1, Age

extCox_LME, Ag, m12						
SEN	SPE	ACC	ROC	Frecuency		
69.2(67.8 70.5)	71.0(70.3 71.6)	70.4(69.8 71.0)	78.6(77.9 79.3)	744	LNS,SDMT,JLO,INMEDIATE,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1	
68.3(66.9 69.6)	71.0(70.3 71.7)	70.3(69.7 70.9)	78.4(77.7 79.2)	724	LNS,SDMT,INMEDIATE,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1	
69.3(68.2 70.5)	70.2(69.7 70.8)	69.8(69.3 70.3)	78.2(77.6 78.8)	1029	LNS,SDMT,HVLTRDLY,SFT,MEANCAU,ABetaTAU,UPDRS1	
68.3(67.2 69.4)	71.5(70.9 72.1)	70.6(70.1 71.1)	78.0(77.4 78.6)	1154	LNS,SDMT,INMEDIATE,HVLTRDLY,SFT,SCOPA,MEANCAU,ABETA,ABetaTAU,UPDRS1	
68.0(66.8 69.3)	70.8(70.2 71.5)	70.0(69.4 70.6)	77.9(77.2 78.5)	896	LNS,SDMT,INMEDIATE,HVLTRDLY,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1	

Tabla 5.8: Resultados de modelos predictivos con análisis de supervivencia de Cox extendido para la población con el criterio SITE al primer año. Utilización de parámetros neuropsicológicos, del DAT-SCAN, CSF y motóricos.

d. Modelos predictivos multivariantes a los dos años del inicio del estudio

Por último, en el segundo año (mes 24), los resultados mejoran con respecto a los dos casos anteriores. Se ve una sensibilidad entre el 73.7 % y 76.2 % y una especificidad entre el 71.5 % y 73.2 %, pero con una menor frecuencia máxima de 876. Se ha aplicado, de igual forma, el Cox extendido y todos los vectores se encuentran

entre 6 y 10 elementos. Se establece el vector de características que se comparte en todos los vectores escogidos y que son:

LNS, SDMT, SFT, MEANCAU, Age

extCox.LME, Ag, m24					
SEN	SPE	ACC	ROC	Frecuency	
75.7(74.5 76.9)	71.8(71.1 72.4)	72.5(72.0 73.1)	81.7(81.0 82.3)	820	LNS,SDMT,,JLO,HVLTRDLY,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1
76.2(74.9 77.5)	71.6(70.9 72.3)	72.4(71.8 73.0)	81.5(80.7 82.2)	705	LNS,SDMT,SFT,MEANCAU,ABetaTAU,UPDRS1
76.1(75.0 77.3)	71.5(70.9 72.2)	72.4(71.8 72.9)	81.5(80.8 82.1)	876	LNS,SDMT,HVLTRDLY,SFT,SCOPA,MEANCAU,ABETA
75.2(73.8 76.5)	72.0(71.2 72.7)	72.5(71.8 73.1)	81.4(80.7 82.1)	679	LNS,SDMT,HVLTRDLY,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1
73.7(72.3 75.1)	73.2(72.5 73.9)	73.1(72.4 73.7)	81.1(80.4 81.9)	753	MOCA,LNS,SDMT,,JLO,INMEDIATE,SFT,SCOPA,MEANCAU,ABETA,ABetaTAU

Tabla 5.9: Resultados de modelos predictivos con análisis de supervivencia de Cox extendido para la población con el criterio SITE al segundo año. Utilización de parámetros neuropsicológicos, del DAT-SCAN, CSF y motóricos.

e. Modelos predictivos multivariantes con todas las medidas al inicio del estudio

Por último, al inicio del estudio, se incluyen todas las medidas para generar los vectores de características de los modelos predictivos. Con esto, se procede a analizar la mejora de añadir otro tipo de características a la generación de estos modelos.

En la tabla 5.10, se pueden observar los vectores de características generados con todas las medidas posibles, análisis Cox y la edad como covariable.

También, para mostrar la efectividad de los modelos progresivos multivariantes y su potencial frente a los modelos progresivos univariantes, se ejecuta el algoritmo 9 al inicio del estudio con la edad como covariable y con el análisis de supervivencia Cox. Así se han obtenido los cinco primeros vectores de características. El último vector es el que se ha establecido en este estudio al analizar las características comunes en los cuatro primeros vectores anteriores.

Al inicio del estudio, los resultados tienen un buen diagnóstico. Se observa una sensibilidad entre el 75.3 % y 76.9 % y una especificidad entre el 70.6 % y 71.4 %. Todos los vectores de características se encuentran entre 7 y 8 elementos. Se establece el vector de características que se comparte en todos los vectores escogidos y que son:

LNS, SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age

En ellos se puede observar como tanto la sensibilidad, como la especificidad, la precisión y ROC han aumentado. Se observan sensibilidades superiores a 73.3 % y especificidades superiores al 68.8 % (valor no superado en los modelos univariantes).

Con estos resultados, se puede hacer visible como los modelos predictivos multivariantes ofrecen una mejor respuesta de predicción que los modelos que únicamente utilizan una variable.

En el último vector de características, propuesto en este estudio, se puede observar la importancia que aportan todas las variables a la predicción. En este último

caso, se encuentran la mayoría de todas las variables entre el 30 y 40 %, exceptuando HVLTRDLY y $A\beta$, que como en los anteriores vectores, tampoco destaca en su aportación al modelo predictivo.

Este último vector, es el que se procede a probar en la realización de los modelos de progresión para comprobar el mejor rendimiento y ver si además, supera los vectores hasta ahora realizados en estudios y trabajos anteriores ($R=0.53$).

Cox, Ag, bl					
SEN	SPE	ACC	ROC	Frequency	
76.1(75.1 77.0)	71.4(70.9 72.0)	72.5(72.0 72.9)	82.5(81.9 83.0)	786	LNS,SDMT,HVLTRDLY,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1
76.9(75.9 77.9)	71.3(70.7 71.9)	72.5(72.1 73.0)	82.4(81.8 82.9)	710	LNS,SDMT,JLO,HVLTRDLY,SFT,MEANCAU,ABETA,ABetaTAU,UPDRS1
76.3(75.6 77.0)	70.6(70.2 71.0)	71.9(71.6 72.3)	81.5(81.2 81.9)	1439	LNS,SDMT,JLO,HVLTRDLY,SFT,SCOPA,MEANCAU,ABETA,ABetaTAU
75.3(74.6 76.0)	70.6(70.3 71.0)	71.7(71.4 72.0)	81.5(81.1 81.8)	1738	LNS,SDMT,HVLTRDLY,SFT,SCOPA,MEANCAU,ABETA,ABetaTAU
76.5(75.6 77.4)	70.6(70.1 71.1)	71.9(71.5 72.3)	81.5(81.0 82.0)	1009	LNS,SDMT,HVLTRDLY,SFT,SCOPA,MEANCAU,ABETA

Tabla 5.10: Resultado del análisis de modelos predictivos multivariantes al inicio del estudio con el análisis Cox y la edad como covariable.

5.3.2. Realización de los modelos de progresión. GRACE

En esta sección, se procede a analizar los resultados proporcionados por los modelos de progresión generados a partir de los modelos predictivos y del análisis masivo realizado en Magerit con la combinación de características.

a. Modelo de progresión según los modelos predictivos multivariantes

Finalmente se realiza una elaboración de un modelo progresión temporal de la enfermedad a partir del paquete en R, GRACE.

Primero se estudian los vectores de características facilitados por los modelos predictivos obtenidos que aporten mejores resultados a los trabajos anteriores ($R=0.53$).

En la tabla 5.11 se presentan los resultados obtenidos para los distintos vectores de características. En ella se recogen el tiempo umbral, la correlación y el MAE.

Modelo	Tiempo umbral (años)	Correlación	MAE	Marcadores
Cox_bl	5.7	0.54	3.69	LNS SDMT HVLTRDLY SFT MEANCAU ABETA Age
extCox_m12	5.4	0.49	3.18	LNS SDMT SFT MEANCAU ABetaTAU UPDRS1 Age
extCox_m24	6.1	0.51	4.08	LNS SDMT SFT MEANCAU Age

Tabla 5.11: Resultados de prueba GRACE para criterio SITE.

El vector seleccionado como más óptimo es el formado por los biomarcadores:

[LNS SDMT HVLTRDLY SFT MEANCAU ABETA Age]

Se selecciona este vector por poseer la correlación más alta, y además superada ($R=0.54$) la que existía hasta el momento en trabajos anteriores ($R=0.53$) y un buen valor MAE⁵.

⁵Medium Absolute Error

Entre los test neuropsicológicos que aportan correlaciones más altas se encuentran:

- HVLTRDLY: *Hopkins Verbal Learning Test Delayed Recall*.
- SFT: *Semantic Fluency Total Score*.

También se encuentra la edad y el ratio de proteínas $A\beta$ y pTau (ABetaTau).

Las trayectorias trazadas por el conjunto de biomarcadores más óptimo puede apreciarse en la figura 5.8.

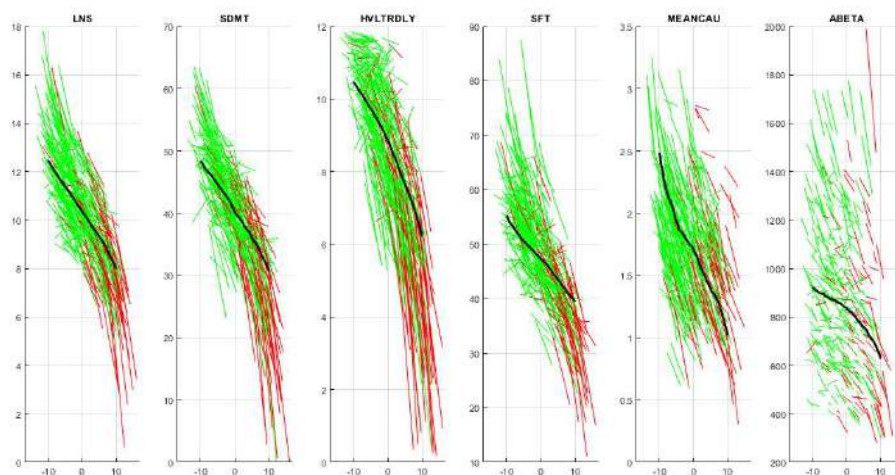


Figura 5.8: Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color rojo para trayectoria short-term de sujetos PD-MCI+, verde para trayectoria short-term de sujetos PD-MCI- y negro para trayectoria long-term.

En las figuras 5.9 y 5.10 se muestran graficadas las matrices de correlación para la pendiente e intercepto, respectivamente. La correlación más elevada se obtiene en la matriz del intercepto, en concreto, se trata de la existente entre las medidas de test SFT y HVLTRDLY con valor de 0.6.

b. Modelo de progresión según el método de combinaciones en Magerit

Se ha ejecutado el *script* en Magerit para comprobar todas las combinaciones de características en vectores de 6 elementos y así, encontrar el vector que tenga un mayor rendimiento.

Por ello se han realizado 51875 combinaciones posibles. Con este código fuente se han obtenido los resultados que se pueden observar en el *.xlsx* que se encuentra en */home/jpalomino/demo/multi_grace_PPMI/*. En esa combinación, el vector de características con mejor resultado ha sido:

SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age

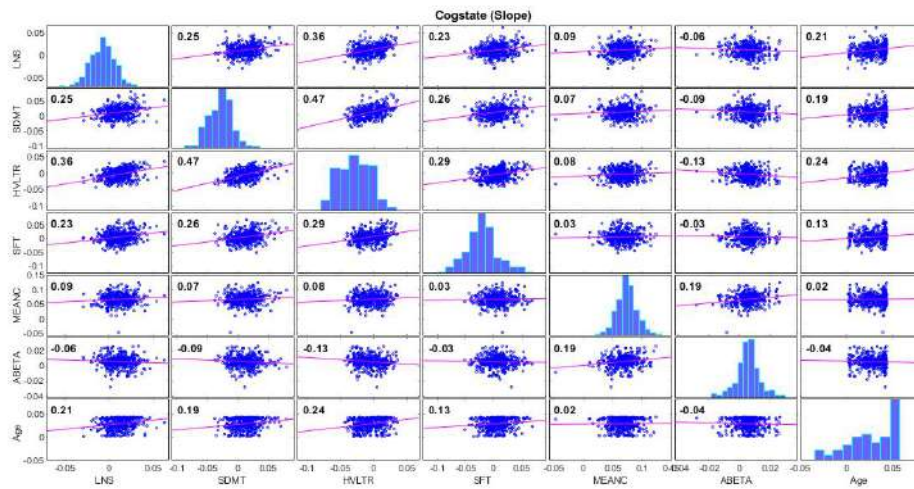


Figura 5.9: Matriz de correlación de la pendiente según el criterio SITE. Se evalúa el vector [LNS SDMT HVLTRDLY SFT MEANCAU ABETA Age]

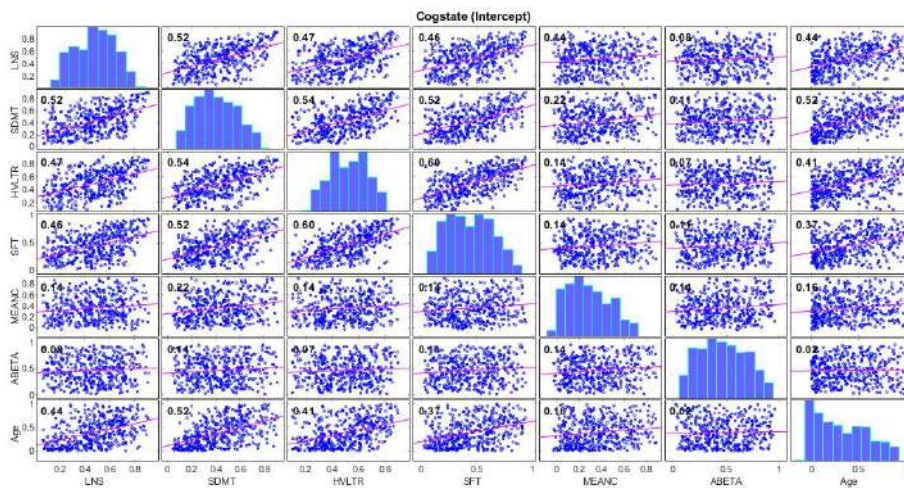


Figura 5.10: Matriz de correlación del intercepto según el criterio SITE. Se evalúa el vector [LNS SDMT HVLTRDLY SFT MEANCAU ABETA Age]

Este vector, además de poseer una correlación de $R=0.5423$, sitúa al 80 % de los sujetos PD-CU antes del tiempo de *onset* (5.4 años). Además el 63 % de la población PD-MCI se encuentra antes y 63 % después del tiempo de *onset*.

En la tabla 5.12, se puede observar los cinco vectores de características que poseen los mejores rendimientos y resultados de todos los analizados en Magerit.

Por lo tanto, este vector de mayor correlación es el aplicado para realizar la historia natural y así validar las trayectorias *long-term*.

Tiempo umbral (años)	Correlación	Marcadores
5.4	0.5423	SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age
5	0.5369	LNS, HVLTRDLY, SFT, MEANCAU, ABETA, Age
5.4	0.5346	SDMT, HVLTRDLY, SFT, ABETA, pTauAB, Age
4.6	0.5341	SDMT, HVLTRDLY, SFT, MEANCAU, pTauAB, Age
5	0.5318	UPSIT, HVLTRDLY, SFT, MEANCAU, ABETA, Age

Tabla 5.12: Resultados del modelo de combinatoria para la obtención de modelos de progresión GRACE en Magerit

Las trayectorias trazadas por el conjunto de biomarcadores más óptimo puede apreciarse en la figura 5.11.

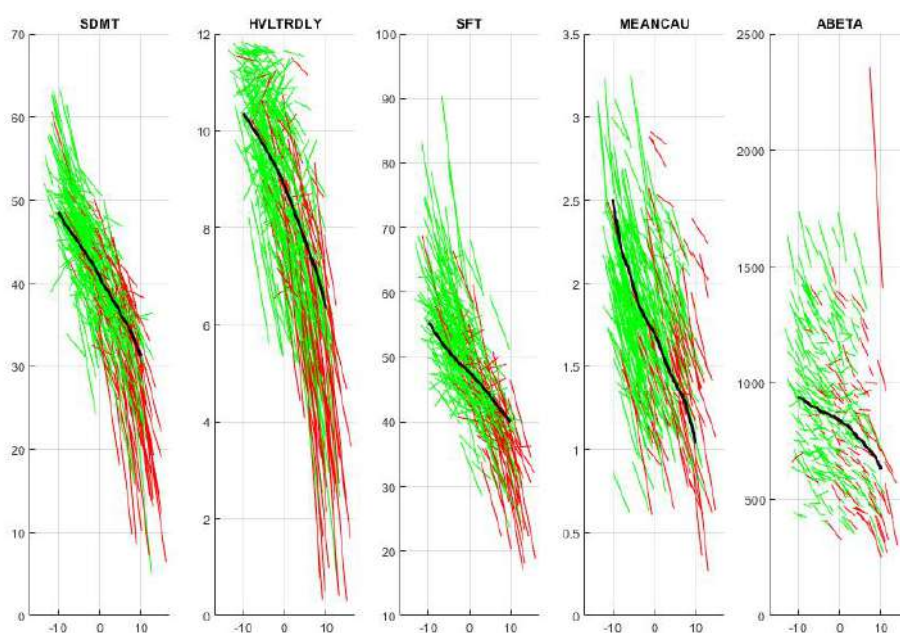


Figura 5.11: Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color rojo para trayectoria short-term de sujetos PD-MCI+, verde para trayectoria short-term de sujetos PD-MCI- y negro para trayectoria long-term.

En las figuras 5.12 y 5.13 se muestran graficadas las matrices de correlación para la pendiente e intercepto, respectivamente. La correlación más elevada se obtiene en la matriz del intercepto, en concreto, se trata de la existente entre las medidas de test SFT y HVLTRDLY con valor de 0.6.

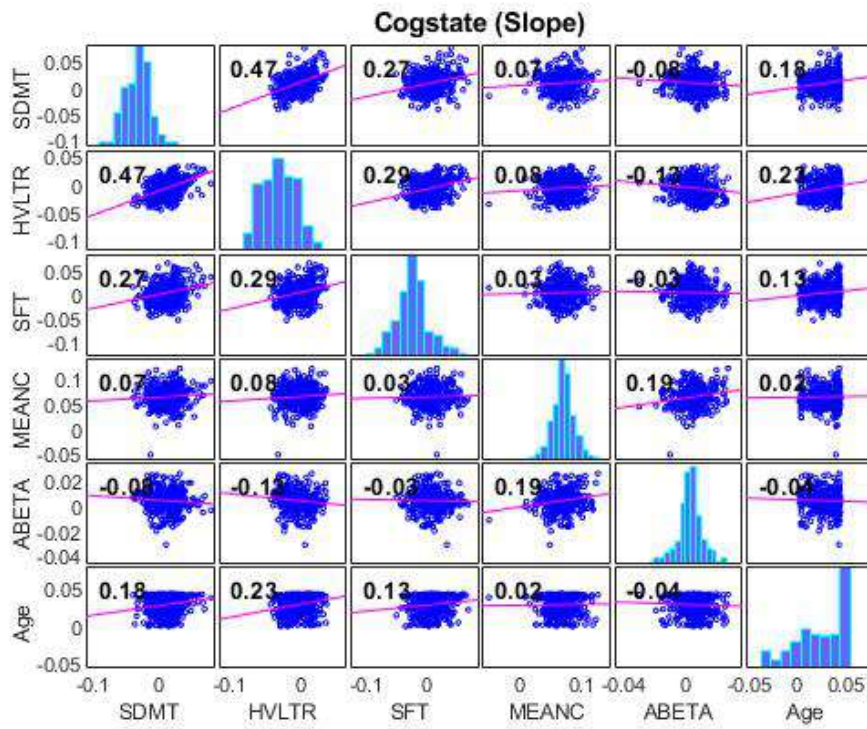


Figura 5.12: Matriz de correlación de la pendiente según el criterio SITE. Se evalúa el vector [SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age]

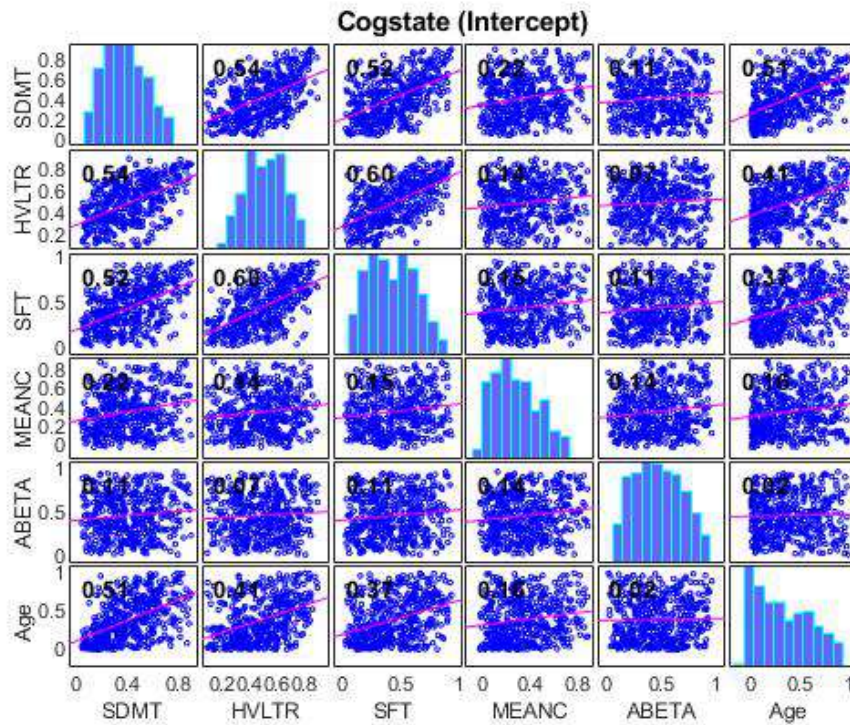


Figura 5.13: Matriz de correlación del intercepto según el criterio SITE. Se evalúa el vector [SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age]

Capítulo 6

Discusiones

En este último capítulo, se realiza un análisis de los resultados obtenidos y mostrados en el capítulo 5 y que se han conseguido en todo el trabajo realizado.

Para el desarrollo de este trabajo se requiere en primer lugar, del estudio de trabajos y artículos previos para el acercamiento y familiarización con la neurociencia, las enfermedades neurodegenerativas y con la enfermedad de Parkinson en particular.

Además de este estudio, es necesario el conocimiento en el uso de herramientas de acceso remoto a máquina o conocimiento estadístico para la elaboración de modelos predictivos y de progresión con GRACE.

6.1. Discusiones

En esta sección, se procede en primer lugar, a discutir y a concluir con las diferencias entre las poblaciones establecidas en los sujetos PD según su estado cognitivo.

En segundo lugar, se analizarán los perfiles AT que se encuentran en la población PD para relacionar estos estados cognitivos con los constructos biológicos de la enfermedad de Alzheimer.

Con los análisis anteriores, se procede a discutir y concluir los factores de riesgo que afectan al progreso y desarrollo de la enfermedad.

Tras este análisis, se obtienen las conclusiones de los modelos predictivos establecidos de manera univariable o multivariable con algunas o todas las características posibles.

Finalmente, se concluye con la historia natural del vector de características óptimo y que, por lo tanto, predice el progreso mediante la historia natural con la validación de las trayectorias *long-term*.

6.1.1. Población PD-CU, PD-MCI y PDD al inicio del estudio

Se ha comenzado con la realización de un análisis de la población al inicio del estudio (*baseline*) de forma transversal, es decir en un instante de tiempo determinado. Este análisis y sus resultados permiten obtener una visión sobre cómo se distribuyen los valores de los diferentes biomarcadores según al grupo clínico a que pertenecen y por tanto, ver si aportan significado o discrepancias en el análisis de los diferentes grupos.

En primer lugar, se han diferenciado los sujetos en dos grupos clínicos: con la enfermedad de Parkinson (PD) y sujetos control o sanos (HC). Entre estos dos grupos clínicos se han observado diferencias significativas en la mayoría de resultados cognitivos, salvo en 3 marcadores (LNS, JLO y ADL).

Además, hay otros marcadores, como es en el caso de $A\beta$, aSyn, Tau y pTau, que también ofrecen valores diferentes entre sujetos PD y HC. Sin embargo, no se presentan diferencias en los convertidores APOE4.

Otras de las conclusiones obtenidas en este primer análisis es la fuerte correlación entre aSyn y pTau y entre pTau y Tau.

Tras esta primera segmentación de los sujetos entre estos dos grupos clínicos, se ha realizado la separación de los sujetos PD en tres grupos según su estado cognitivo: sin declive cognitivo (PD-CU), con declive cognitivo leve (PD-MCI) y con demencia (PDD).

Entre estos tres grupos, se ha vuelto a realizar un análisis de sus características al inicio del estudio para que al igual que en el caso anterior entre sujetos PD y HC, se puedan obtener discrepancias y establecer factores de riesgo y características que permitan determinar el estado cognitivo del paciente.

En este caso, se obtienen diferencias significativas en las características que se muestran en la tabla 6.1:

Características demográficas y clínicas				
Sujetos	323 (PD-CU)	83 (PD-MCI)	17 (PD-Dementia)	Multicompare
Mujeres	120 (37.2%)	20 (24.1%)	6 (35.3%)	
Edad	60.13 (9.82) (33.50,82.259)	66.01 (7.37) (48.03,84.88)	69.45 (7.45) (54.82,82.98)	
Resultados cognitivos				
SDMT	42.94 (8.75) (17.00 82.00)	36.69 (10.18) (7.00 58.00)	29.82 (10.85) (15.00 50.00)	abc
Marcadores de CSF e imágenes de DAT				
ABETA	942.40 (413.12) (315.90 3707.00)	844.08 (395.26) (238.80 2572.00)	592.27 (251.10) (308.80 1300.00)	c
TAU	167.17 (54.61) (80.93 467.00)	172.53 (59.27) (83.93 345.30)	202.68 (78.07) (97.17 327.40)	c
PTAU	14.52 (4.99) (8.02 40.13)	15.51 (5.40) (8.36 32.80)	18.59 (7.68) (8.01 30.84)	c
Abeta/TAU	5.78 (1.52) (1.38 10.38)	5.14 (1.54) (1.19 8.08)	3.24 (1.38) (1.39 6.30)	abc

Tabla 6.1: Características demográficas y clínicas con más discrepancias entre sujetos PD-CU, PD-MCI y PDD a los 5 años de seguimiento con el criterio SITE. Los datos son la media (desviación estándar) (mínimo, máximo) o n (%). a: discrepancia entre PD-CU y PD-MCI; b: discrepancia entre PD-MCI y PDD; c: discrepancia entre PD-CU y PDD

Con esa tabla, se reflejan las discrepancias existentes entre los tres grupos clínicos según su estado cognitivo y que coinciden con los resultados de los artículos que

también hacen análisis al inicio del estudio y que realizan separación entre estos tres grupos [35] [53]. Entre esos resultados, se encuentra las diferencias existentes en el sexo. Este factor se considera, de igual forma, importante y a considerar en el artículo de Chorlerton et al. [5]. Por ello, este estudio está de acuerdo con esa conclusión, ya que establece un 64.7% de población masculina en el grupo PDD (figuras 6.2 y 6.1).

Otro de las medidas que se enfatiza en los artículos de estudio ha sido MoCA. El *cutoff* de esta medida depende del autor propuesto. Aún así, en este caso de estudio no se muestran unas diferencias muy altas en esa medida.

Sin embargo, hay otras medidas, como el caso de marcadores CSF y procedentes de imágenes DAT que si que presentan diferencias en coincidencia con autores. Este es el caso de $A\beta$, que tiene efecto decreciente según se produce el avance desde PD-CU hasta PDD.

También, entre este tipo de marcadores, se encuentran pTau y Tau, por ejemplo, pero con efecto ascendente en la evolución del estado cognitivo hacia demencia.

Terminado el análisis según la observación de los resultados del presente estudio y su correspondencia con los artículos, se ha realizado un análisis de post-procesado de multi-comparación entre estos tres grupos. Toda la información detallada y con mayor profundidad puede encontrarse en el apéndice B.

Con este análisis, se han obtenido otras características que discrepan entre ellas y que no se han podido obtener con un análisis de la tabla 5.3 sin el análisis de multi-comparación.

Así, se añade como discrepancia y factor de riesgo a la medida SDMT, ya que presenta discrepancia entre los tres grupos cognitivos. Otro de los marcadores que se incluye en procedentes de imágenes DAT y marcadores de CSF es el ratio $A\beta$ /Tau. Por lo tanto esta relación también puede considerarse como factor de riesgo para futuras aportaciones al trabajo.

Sin embargo, si se quieren obtener discrepancias entre sujetos sin declives con sujetos tanto PD-MCI como PDD, se encuentra con otra clase de marcadores. Estas, son las medidas cognitivas, a excepción de ADL. También al igual que este caso, se añade la medida del caudado (MEANCAU).

Por ello, es lógico y coherente que en los modelos predictivos de la enfermedad como factores importantes que predicen el paso de estado cognitivo, se encuentren SDMT, $A\beta$ /Tau y la edad.

También, para finalizar la discusión entre los grupos, se ha analizado la aportación de la patología amiloide a los sujetos. Otros artículos también hablan de esta patología como es en el caso de Fiorenzato et. al [16] e Irwin et. al [8]. En este caso de estudio, se ha establecido que un tercio de los sujetos PD poseen patología amiloide ($A\beta < 683.45$ pg/ml). Si se realiza distinción entre los sujetos que poseen

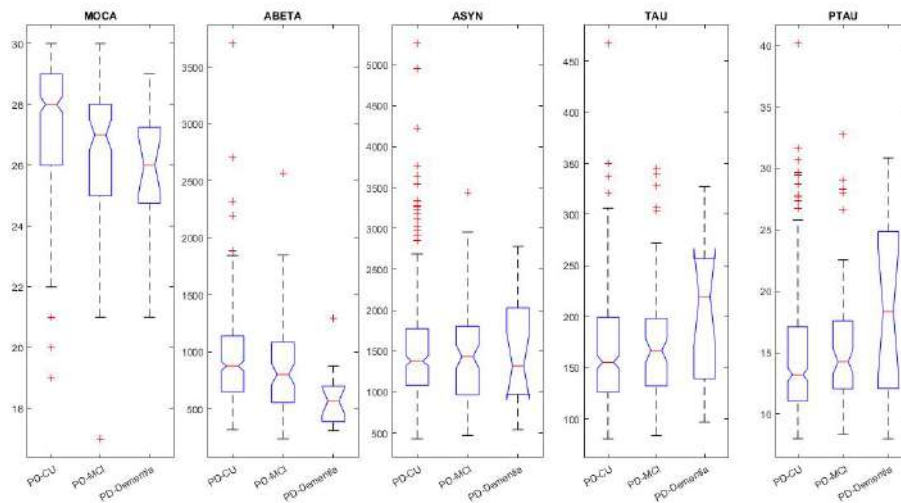


Figura 6.1: Valores al inicio del estudio entre sujetos PD-CU, PD-MCI y PDD.

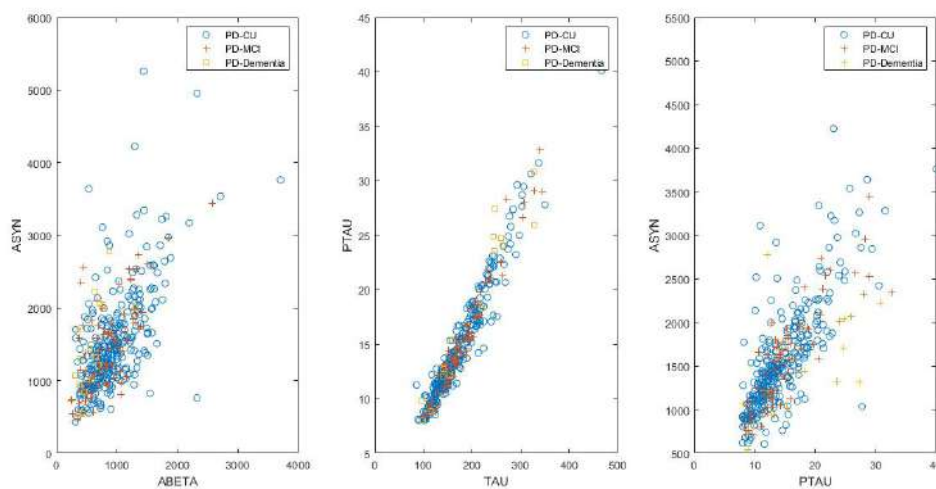


Figura 6.2: Correlación entre marcadores de CSF y aSyn al inicio del estudio entre PD-CU, PD-MCI y PDD.

esta patología y los que no, hay diferencias significativas entre $A\beta$, aSyn, Tau y pTau.

Sin embargo, en este caso, no hay diferencias en las medidas cognitivas según su patología amiloide resultante.

6.1.2. Perfiles AT en sujetos PD

La última acción realizada, es la de analizar los resultados de la patología amiloide en los sujetos de la población. Con esta separación, se ha procedido a relacionar el estado cognitivo de los sujetos con la enfermedad de Parkinson con los constructos

biológicos de la enfermedad de Alzheimer.

Esta relación se ha establecido mediante el perfil AT. Este perfil AT(N) se analiza en Jack et. al [6]. Se define A+ cuando $A\beta < 685.45$ pg/ml y T+ cuando $p\text{Tau} > 23$ pg/ml. En este caso de estudio, se encuentra que el perfil A+ es del casi un tercio de la población PD-MCI (26.5%). De todos modos, esta patología que se ha descrito anteriormente aumenta desde el grupo de sujetos PD-CU (23.4%) hasta PDD (66.7%), donde es 2/3 de ese grupo.

Las conclusiones anteriores resultan del análisis al inicio del estudio. En cambio, en el análisis longitudinal, estos porcentajes aumentan en todos los grupos, excepto en PDD, donde se mantiene constante.

También, a ese perfil A se le puede adjuntar el perfil T. Con estos dos perfiles se obtiene que 2/3 de la población PD-MCI es parte del perfil A-T-. Sin embargo, 2/3 de los sujetos PDD se encuentran en Alzheimer continuo (A+). Con esto se concluye en que 1/3 de los sujetos PDD son A+T+ y más de la mitad son A+ (66.7%), aproximado al artículo de Jack et al. [6].

Con los resultados obtenidos se ha relacionado la enfermedad de Parkinson con la enfermedad de Alzheimer a través de este perfil AT.

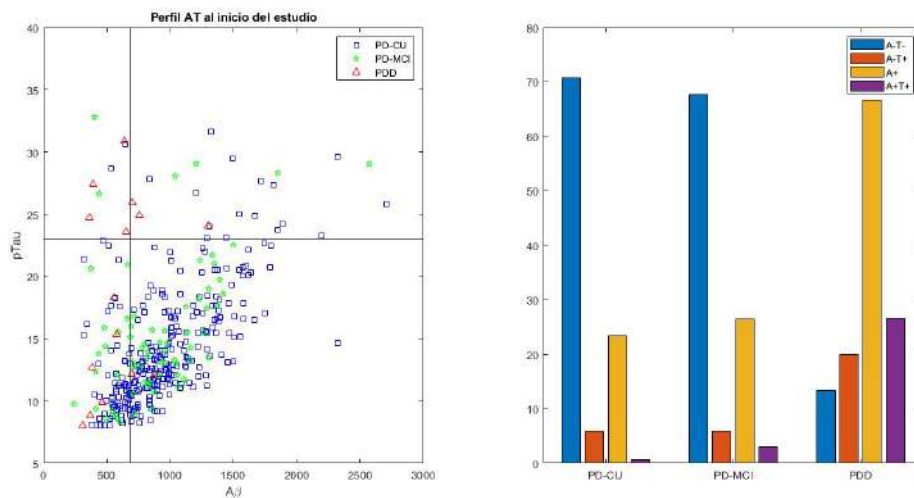


Figura 6.3: Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores al inicio del estudio. Las líneas muestran los puntos de umbral en las medidas.

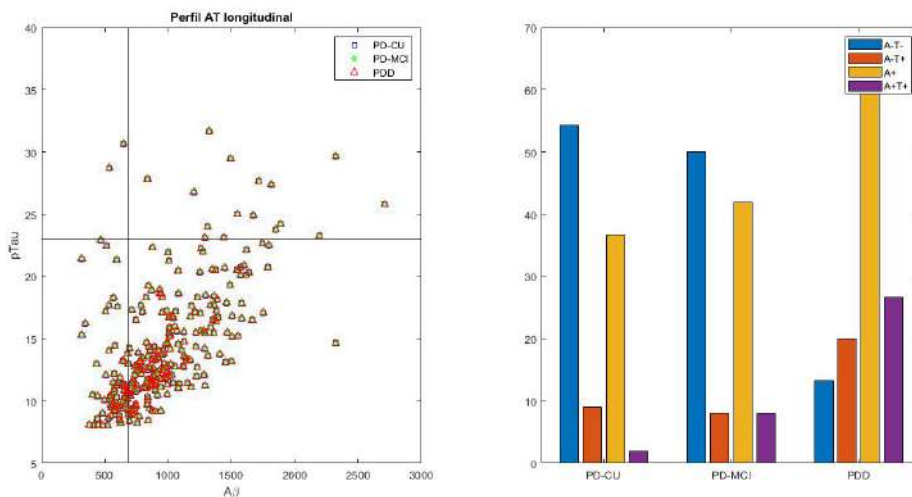


Figura 6.4: Porcentajes de los perfiles AT en sujetos PD-CU, PD-MCI and PDD. El criterio A+ es usado con un $A\beta < 683.45$ pg/ml. El T+ es establecido por $p\text{Tau} > 23$ pg/ml. Corresponde con los valores en el análisis longitudinal. Las líneas muestran los puntos de umbral en las medidas.

6.1.3. Factores de riesgo

Después de observar las diferencias entre los distintos marcadores según sus grupos clínicos y la obtención de características que pueden actuar como predictores y factores de riesgo, se han procedido a analizar todos aquellos que se establecen según los distintos autores y que se analizan en la población de estudio.

También se ha analizado el perfil AT(N) para relacionar estos estados cognitivos con la AD. Por ello, se puede establecer A+ como predictor de la enfermedad [16] [8]. Otro de los marcadores utilizados ha sido APOE4.

Además, a esto se le suma el factor del sexo y que se establece en el artículo de Chorlerton et al. [5]. Con el análisis de artículos, un predictor que se considera útil es el marcador de la fase del sueño REM. Por último, se ha establecido el análisis del PACC definido por Donohue et al. en su artículo [33].

El *cutoff* de MoCA puede poseer diferentes en función del autor que realiza el estudio. En este caso, se ha establecido el valor de $\text{MoCA} < 24$ según Aarsland et. al [7]. El PACC, del mismo modo, se ha establecido con el valor de $\text{PACC} < -1.0$ según [33].

Con esas consideraciones, en esta población se ha establecido que efectivamente MoCA es un predictor para el paso desde no encontrarse con declive a MCI o demencia, pero no diferenciable entre MCI y demencia, por lo que no es del todo el resultado que se esperaba.

Por otra parte, el APOE4 no ofrece resultados visible según se había confirmado en los análisis previos. Sin embargo, la patología A+ si que ofrece ser predictor del

estado cognitivo con un aumento de porcentaje desde PD-CU hasta PDD.

En el caso del sexo, aunque si que se veía la diferencia entre sujetos PD y HC no es tan elevada, ni sigue una tendencia entre los tres grupos de estudio. Del mismo modo, no se ve esa tendencia en el análisis de REM.

Como resultados favorables, en los grupos PD-CU, PD-MCI y PDD, se tiene la patología A+ (23.4 %, 26.5 % y 66.7 %, respectivamente) y PACC<-1.0 (22.91 %, 51.81 % y 70.59 %, respectivamente).

6.1.4. Modelos predictivos al inicio del estudio

De los objetivos que se buscan en este trabajo, es de encontrar aquellos marcadores que permitan diferenciar entre sujetos con Parkinson y deterioro cognitivo (PD-MCI+) y aquellos sin deterioro cognitivo (PD-MCI-).

Para ello, es necesario realizar un análisis longitudinal, ya que sus resultados mejoran a los de un análisis transversal gracias a características como la sensibilidad, especificidad o ROC. Estas variables, entre otras mejoran al ampliar el número de visitas de los sujetos de este estudio.

Los vectores que se han obtenido en la realización de los modelos predictivos al análisis del estudio han sido en gran frecuencia de pruebas cognitivas como LNS, SDMT, SFT o HVLTRDLY. También otros, como JLO o SCOPA.

Además al incluir el resto de características, se pueden encontrar otras de imágenes DAT como la absorción media en el caudado (MEANCAU), pero no del putamen de dopamina (MEANPUT). Aún así, si aparecen medidas del CSF, como es el caso del ratio $A\beta/\text{Tau}$ o $A\beta$.

Otras de las medidas que aparecen, es el UPDRS1, aunque no es de interés en el conjunto de características de estudio.

Bernal et al. [27] [26] trata en sus artículos los LME que son necesarios en la realización de los modelos predictivos y que se usan para la obtención de nuestros vectores de características.

En este estudio se ha hecho énfasis en diferenciar los aquellos modelos predictivos que dependen de una sola variable de los que dependen de un grupo de variables no motóricas y de los que dependen de todo tipo de variables. Todos estos análisis implican a la edad como covariable debida a su influencia.

Tsiouris et al. [56], Smith et al [53], Tang et al. [54] y Platero et. al [63] [42] tratan en sus artículos la realización de estos modelos predictivos, la importancia de la aportación de medidas no únicamente no motóricas a los modelos predictivos y la diferencia de realizar modelos predictivos univariantes o multivariantes.

En primer lugar, se han analizado todos los modelos predictivos univariantes.

Con todos esos modelos, se han destacado REM, LNS, SDMT, INMEDIATE, SFT y MEANCAU. Estos modelos univariantes tienen valores adecuados para ser incluso de una variable. Aún así, se realiza una mejora únicamente con medidas cognitivas.

Al añadir medidas cognitivas, el resultado mejora en cuanto a la precisión y el valor de ROC, próximos al 80 %. Con esto se demuestra que añadiendo las medidas cognitivas a los modelos predictivos, el resultado de predecir a los sujetos PD-MCI es mayor.

En este caso, el vector de características resultante, es el formado por LNS, SDMT, JLO, SFT y SCOPA con la edad como covariable (SEN=73.0 %, SPE=68.8 %, ACC=69.8 %, ROC=77.7 %, Frequency= 2640).

Aún así, se pueden mejorar estos modelos añadiendo todas las posibles medidas de nuestros sujetos. Si se añaden todas las características, las medias de sensibilidad suben del 75 % y la especificidad se encuentra por encima del 70.6 %. De esa forma, se establece que el mejor vector de características es el formado por LNS, SDMT, HVLTRDLY, SFT, MEANCAU, ABETA junto con la edad como en el caso anterior (SEN=76.5 %, SPE=70.6 %, ACC=71.9 %, ROC=81.5 % y Frequency=1009).

Así, de acuerdo con Tang et al. [54], se comprueba la mejora de obtener un vector de características más válido para la predicción. Con este vector, si no se hubiera desarrollado un nuevo método de combinatoria, hubiera sido el elegido en los modelos de progresión de GRACE [34] y comprobado mediante la historia natural de los valores de los marcadores en el *long-term*.

6.1.5. Historia natural mediante los modelos de progresión. GRACE

Modelar la enfermedad de Parkinson y su progresión con el uso de GRACE, permite ordenar en el tiempo a los sujetos de diferente estado cognitivo. Con este modelo de progresión, se modela el proceso continuo de la enfermedad de Parkinson con marcadores que progresan también de forma continua.

Gracias al artículo de Schrag et al. [2], se pueden establecer los valores que tendría los marcadores de los sujetos PD-MCI, junto con otros como puede ser MoCA en el trabajo de Aarsland et al. [7].

En primer lugar, se ha realizado la progresión de la enfermedad a partir del modelo predictivo que se ha generado anteriormente. Con este modelo, se ha obtenido una correlación mayor que en trabajos anteriores con $R=0.54$. Las trayectorias se pueden observar en la figura 6.5.

Tras este análisis según lo establecido por los modelos predictivos, a través del método de combinatoria basado en el *script* de R. Batal (Algoritmo 12) se han realizado 51875 combinaciones distintas de 5 elementos de todas las características posibles.

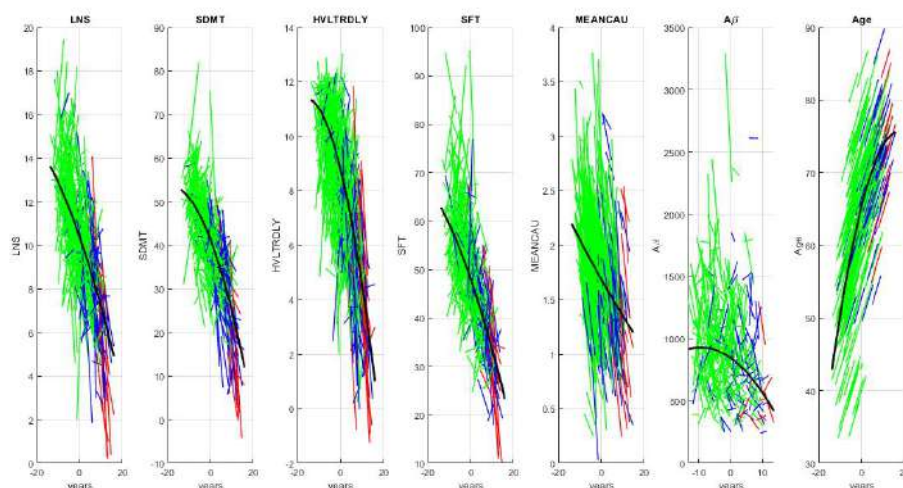


Figura 6.5: Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color verde para trayectoria short-term de sujetos PD-CU, color azul para trayectoria short-term de sujetos PD-MCI, color rojo para trayectoria short-term de sujetos PDD y negro para trayectoria long-term del modelo propuesto.

En este caso, el vector seleccionado ha sido el formado por SDMT, HVLTRDLY, SFT, MEANCAU, ABETA y la edad como covariable. Con este modelo de características se ha obtenido una correlación de $R=0.5423$. Este conjunto de características, presenta un tiempo de *onset* de 5.4 años. Con estos datos, se ha procedido a estimar el *long-term* de los marcadores de interés definida la historia natural.

Las trayectorias se pueden observar en la figura 6.6.

Se procede a analizar la validación de los resultados en la tabla 6.2.

Características clínicas y demográficas					Modelo
Sujetos	PD-CU	PD-MCI	PDD	clinical cutoff	$t_{onset}=5.4$ years
Edad	60.13	66.01	69.45	66.1 [2]	67.88
GDS	2.17	2.86	2.71	2.0 [2]	2.18
Resultados cognitivos					
MOCA	27.33	26.58	25.94	26 [7]	25.28
Marcadores de CSF e imágenes DAT					
ABETA	942.40	844.08	592.27	310.5 [2]	714.69
ASYN	1519.17	1470.97	1443.49	1753.1 [2]	1281.32
A β /TAU	5.78	5.14	3.24	7.8 [2]	4.88
MEANCAU	2.07	1.80	1.59	1.8 [2]	1.49

Tabla 6.2: Para cada marcador seleccionado, una comparación entre los valores medios por grupo clínico (PD-CU, PD-MCI, PDD) al inicio, con respecto a los valores estimados por las trayectorias long-term al inicio de la demencia y el cutoff que se utiliza en la práctica clínica para determinar la progresión a la demencia.

En la tabla 6.2 vemos la correspondencia entre los valores de cada grupo, el valor estimado en la trayectoria long term y los *cutoff* clínicos utilizados para ver la

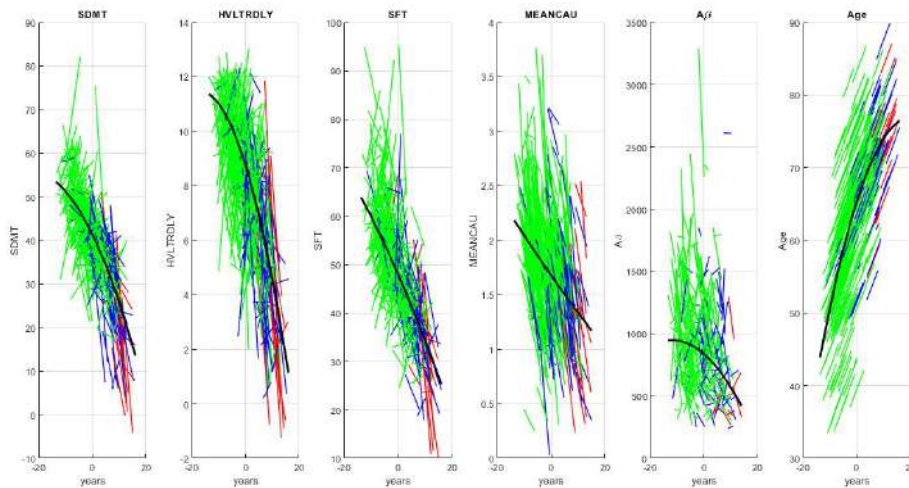


Figura 6.6: Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color verde para trayectoria short-term de sujetos PD-CU, color azul para trayectoria short-term de sujetos PD-MCI, color rojo para trayectoria short-term de sujetos PDD y negro para trayectoria long-term del modelo propuesto.

correspondencia entre el valor clínico y el estimado.

Podemos observar como los marcadores relacionados con las características clínicas que según las secciones anteriores ayudan a predecir la enfermedad, están de acuerdo en la parte del *cutoff* clínico con el modelo propuesto a los 5.4 años.

En la figura 6.7 podemos observar las trayectorias *long-term* de los marcadores que se comprueban en la tabla 6.2. A través de ellas observamos los valores obtenidos desde la curva que representa el modelo propuesto.

También se analiza la característica MoCA, que permite evaluar la cognición global. Además, indica la existencia de deterioro cognitivo al indicarse un valor de $MoCA < 26$ [7] para la conversión a PD-MCI. Al analizar este valor a los 5.4 años, se obtiene un valor de 25.28, es decir, muy próximo al establecido clínicamente.

En cambio, en los valores de los marcadores de CSF y los procedentes de imágenes DAT, encontramos discrepancias entre la población obtenida para nuestro estudio y los valores clínicos obtenidos [2].

Aún así, son valores cercanos, aunque no próximos a los deseados en el modelo que hemos obtenido. Por ejemplo, en el caso de la medida del caudado, el modelo indica un valor de 1.49, cuando el valor clínico se establece en 1.8.

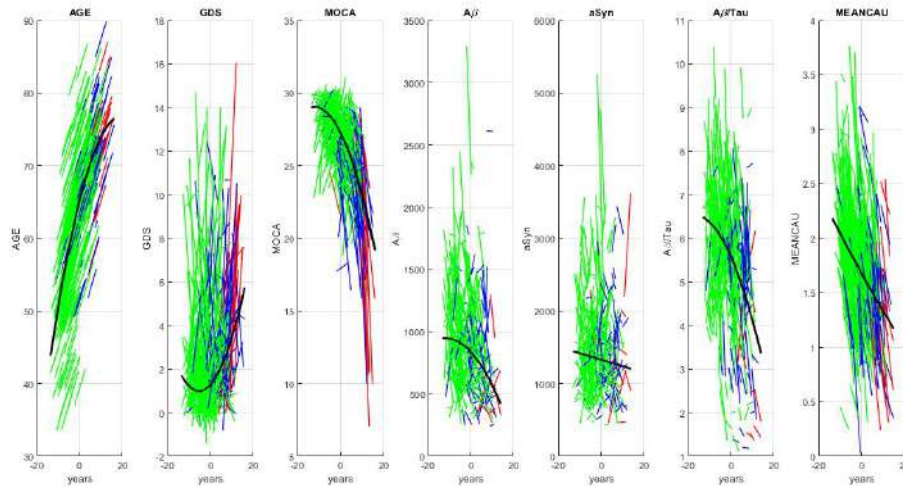


Figura 6.7: Trayectorias long-term y short-term obtenidas a través del criterio SITE. Color verde para trayectoria short-term de sujetos PD-CU, color azul para trayectoria short-term de sujetos PD-MCI, color rojo para trayectoria short-term de sujetos PDD y negro para trayectoria long-term del modelo propuesto.

Long-term en porcentajes

Además, para confirmar la validez del modelo propuesto, se puede observar la figura 6.8. Esta figura muestra las trayectorias long-term de los marcadores del modelo que se ha propuesto en porcentajes (escala común). Se observa que los marcadores HVLTRDLY y ABETA son alrededor de 25 % en los 15 años anteriores al tiempo de *onset*.

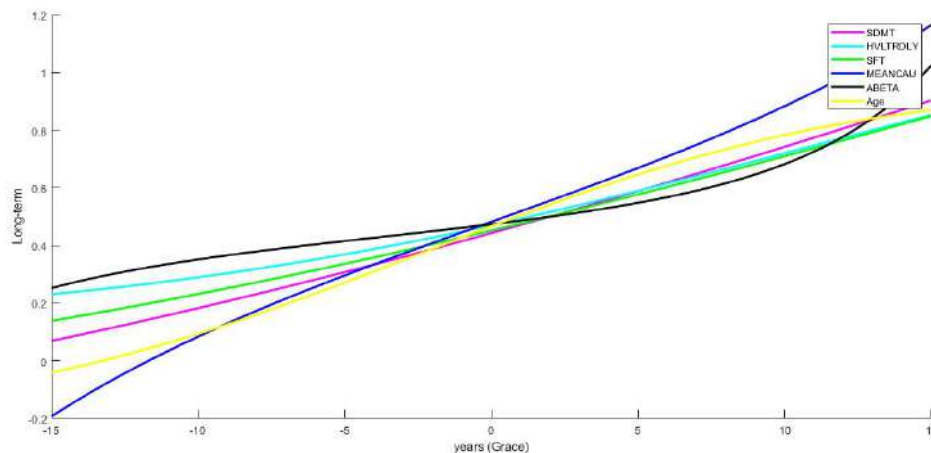


Figura 6.8: Trayectorias long-term de los marcadores del modelo propuesto (SDMT, HVLTRDLY, SFT, MEANCAU, ABETA, Age) en una escala común, i.e en percentiles

Por otro lado, se observa que las medidas del caudado y de la edad están con

valores por debajo del 0% (aproximadamente -20% y -5%, respectivamente). Sin embargo, los marcadores SDMT y SFT se encuentran alrededor de un 10% y un 15%, respectivamente.

A partir de 10 años anteriores al año 0 en GRACE, las pendientes se mantienen de forma aproximadamente iguales en la edad y MEANCAU hasta el año 0. Por otro lado, ocurre lo mismo, pero desde el año 0 en las medidas de SFT, SDMT y HVLTRDLY hasta prácticamente el final del tiempo. Al seguir la misma trayectoria, indica que los marcadores siguen el mismo proceso biológico.

Al finalizar el tiempo, MEANCAU supera el 100%, mientras que el resto de marcadores (HVLTRDLY, SFT, MEANCAU, ABETA, Age) se encuentran cercanos al 80%, menos ABETA, próximo al 100%.

Todo este análisis permite encontrar los marcadores que están presentes en las fase preclínicas de la enfermedad. Al detectarse antes de la propia enfermedad, se puede mejorar la calidad de vida de los sujetos.

6.2. Conclusiones.

Tras la discusión de los resultados obtenidos y la comparación con artículos de diversos autores, se procede a exponer las conclusiones obtenidas.

Respecto al análisis de los grupos PD según su estado cognitivo al inicio del estudio, se han establecido diferencias en primer lugar entre sujetos PD y HC en la mayoría de marcadores excepto LNS, JLO, ADL y APOE4. Además, se obtiene correlación aSyn-pTau y pTau-Tau. Después, en la diferenciación de grupos PD, se observan discrepancias en el sexo, marcadores CSF como $A\beta$, pTau y Tau. Otro resultado cognitivo que se suma en el post-procesado de multicomparación es SDMT. También el ratio $A\beta$ /Tau y la edad como covariable.

En el análisis de perfiles AT en los sujetos PD, el perfil A+ es casi 1/3 de la población PD-MCI y 2/3 en PDD, más del 50%.

Si se habla de los factores de riesgo, se proceden a analizar los citados en los artículos para su comprobación, obteniendo que MoCA no es del todo un buen predictor, APOE4 no ofrece resultados claros. Tampoco es clara la diferencia en el caso del sexo, al igual que en REM. Sí que se observan favorables, la patología A+ y PACC.

Por último, en cuanto a los modelos predictivos y su comprobación y evaluación de la historia natural de los marcadores, se obtiene como vector con mayor rendimiento el compuesto por las características LNS, SDMT, HVLTRDLY, SFT, MEANCAU, ABETA y la edad con una correlación de $R=0.5423$ tras las diferentes combinaciones y con un tiempo de *onset* de 5.4 años. Al compararlo con los resultados clínicos, se aproxima a los establecidos.

6.3. Limitaciones del trabajo.

En este estudio, respecto a los anteriores, se ha reducido el coste computacional al desarrollar otro algoritmo más en Magerit, cosa que no se realizaba en los trabajos de M.Ruiz [48] y M.Lerín [1].

Aunque en estos trabajos anteriores sí que se ha reducido el coste computacional en otros aspectos que ya se han realizado en el presente estudio, la realización de los modelos de progresión y la prueba de combinaciones de características para encontrar el vector de marcadores con el mejor rendimiento se han realizado en este, aumentando hasta un total de 51875 combinaciones posibles.

Aún así, una de las principales limitaciones en este estudio es el número de visitas y la poca trayectoria que tiene PPMI con un máximo de 5 años de estudio. Esto deriva en que los modelos predictivos y sus medidas de sensibilidad o especificidad no lleguen a ser todo lo óptimas que se puede esperar, igual que los resultados en los modelos de progresión.

Se espera que los resultados se vean mejorados en cuanto al número de visitas para la obtención de unos mejores resultados capaces de predecir con un mejor rendimiento el progreso de la enfermedad.

Además, el tiempo de ejecución que se ha tenido para poner en marcha en Magerit el algoritmo capaz de realizar combinaciones ha sido menor que el esperado. Esta limitación puede solventarse con la ejecución de este *script* con un tiempo considerable para obtener un número de combinaciones más elevado.

6.4. Desarrollos futuros

Hay diversos desarrollos futuros que se pueden plantear a partir de este proyecto para su mejora o implantación de nuevas líneas de estudio e investigación:

- Obtener una relación más profunda de la enfermedad de Parkinson con la enfermedad de Alzheimer a parte de la de los estados cognitivos con los constructos biológicos por el perfil ATN.
- Mejorar *scripts* realizados, así como comentarios, notas u otros elementos que permitan que sea más eficiente y útil para nuevos usuarios.
- Seguir aumentando las combinaciones de modelos predictivos para conseguir una batería amplia de vectores posibles.
- Aumentar la población de estudio y sus visitas para obtener más sujetos con visitas en el tiempo.
- Incluir estas visitas y sujetos de otras bases de datos para contrastar resultados, no únicamente de PPMI.

Apéndice A

Análisis estadístico aplicado a ciencias de la salud

A.1. Introducción

La ingeniería y el análisis estadístico son un componente esencial de toda investigación biomédica, y el uso de técnicas estadísticas ha crecido exponencialmente en los últimos años en el campo de la investigación en ciencias de la salud.

El uso de técnicas estadísticas más específicas en investigación se ha incrementado en las últimas décadas, impulsado por la inclusión de la bioestadística en los planes de estudio de los profesionales de la salud y la inclusión de historias clínicas Expertos metodológicos en los grupos de investigación. El análisis estadístico utilizado en el estudio depende en gran medida del tipo de estudio, los objetivos a alcanzar, el tamaño de la muestra, así como el nivel de comprensión de los investigadores de las técnicas estadísticas y el software para realizar la tarea.

Esta es la razón por la cual la estadística juega un papel tan esencial en la investigación en ciencias de la salud, y la investigación de mejor calidad es realizada por un equipo multidisciplinario que incluye profesionales médicos, académicos e investigadores expertos en métodos estadísticos.

El propósito de este apéndice es proporcionar una breve introducción a la prueba y métodos estadísticos útiles para la implementación de este proyecto.

A.2. Conceptos básicos

A.2.1. Inferencia estadística

La inferencia estadística es un conjunto de métodos y técnicas que permiten, a partir de la información empírica proporcionada por una muestra, inferir el comportamiento de una determinada población con un riesgo de error medible en términos de probabilidad.

Los métodos paramétricos para la inferencia estadística se pueden dividir principalmente en dos: métodos de estimación de parámetros y métodos de prueba de

hipótesis. Ambos métodos se basan en el conocimiento teórico de la distribución de probabilidad de una muestra estadística utilizada como estimador de parámetros.

A.2.2. Pruebas de hipótesis

Los métodos de prueba de hipótesis que verifican si una determinada hipótesis referida a un parámetro poblacional, o parámetros similares de dos o más poblaciones, son consistentes con la evidencia empírica dada en la muestra.

Las suposiciones hechas sobre los parámetros se denominan suposiciones de parámetros. Para cualquier hipótesis paramétrica, la prueba depende de la creación de un criterio de decisión, que en cada caso depende de la naturaleza de la población, de la ley de probabilidad del estimador de ese parámetro, y del control que se quiera fijar de antemano.

Es importante entender tres conceptos, para entender el concepto:

- Inferencia: Literalmente, hacer juicios o inferencias a partir de ciertos supuestos, ya sean generales o específicos.
- Población: La cantidad total de datos contenidos en una variable.
- Muestra estadística: una muestra que forma parte de la población de datos.

En todo contraste intervienen dos hipótesis:

Hipótesis nula (H_0)

La hipótesis nula (H_0) es una hipótesis en la que el parámetro determinado (X) toma un cierto valor (X_0) y es una hipótesis que tiene carga de prueba.

Hipótesis alternativa (H_a)

La hipótesis alternativa (H_a) es una hipótesis en el que valor numérico que describe las características del estudio (X) es diferente del valor sugerido por la hipótesis nula (X_0). Este valor puede ser mayor o menor que el valor de la hipótesis nula. La hipótesis alternativa puede ser unilateral o bilateral.

- La hipótesis alternativa unilateral determina si un parámetro de estudio difiere del parámetro propuesto en una dirección. En otras palabras, solo puede ocurrir uno de los siguientes:
 1. $X < X_0$
 2. $X > X_0$
- Si hablamos de hipótesis alternativa bilateral, esta busca esclarecer si el valor del estudio es mayor o menor al sugerido. En otras palabras, solo puede administrar el caso $X \neq X_0$.

En las pruebas de significancia estadística, las pruebas de una y dos colas son formas alternativas de calcular la significancia estadística de un parámetro inferido de un conjunto de datos.

Una prueba de dos colas es apropiada si el valor estimado está por encima o por debajo de cierto rango, mientras que una prueba de una cola es apropiada si el valor estimado puede desviarse del valor de referencia en una sola dirección, izquierda o derecha, pero no en ambas (Figura ??)

A.2.3. Estadísticos y p-valor

Los *estadísticos* son el resultado del contraste de hipótesis. Un estadístico es un parámetro empírico cuyo valor proviene de datos de muestra y sigue una distribución de probabilidad específica.

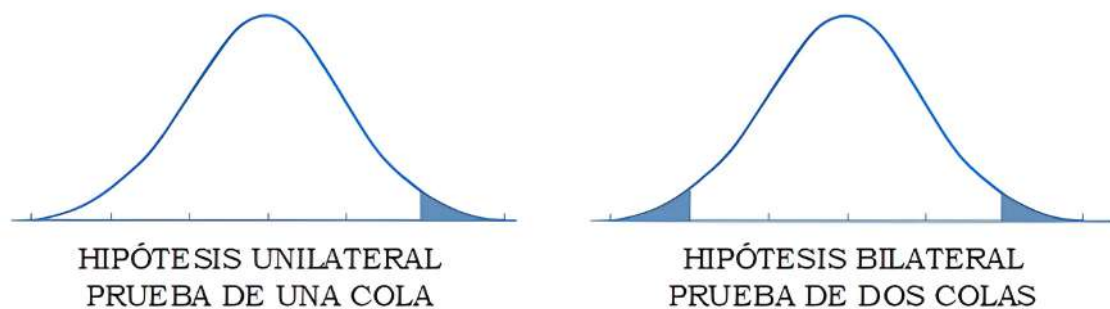


Figura A.1: Correspondencia entre hipótesis unilateral y prueba de una cola e hipótesis bilateral y prueba de dos colas

Se puede hablar de la probabilidad que tiene el estadístico de alcanzar un valor determinado debido a que ese valor estadístico depende de una determinada distribución de probabilidad. Para referirnos a esta probabilidad, hablamos de *p-valor*. Como probabilidad que es, varía entre cero y uno. Se parte desde la suposición de que la hipótesis nula es verdad, haciendo que los p-valores muy pequeños darán un suceso muy poco probable y viceversa.

Debido a las distintas formas de la entrada de los datos (paramétricos, cualitativos, continuos, discretos, etc.), existen estadísticos especializados en el tratamiento de ese tipo concreto de datos. Se pueden destacar la *distribución normal Z*, *T de Student* o *F de Snedecor*.

Para establecer si una hipótesis es verdadera o no, se establecen las *regiones de rechazo*. Estas, son intervalos de la distribución en los que la hipótesis nula es falsa. Si el p-valor del estadístico está dentro de la región de rechazo, la hipótesis nula se rechaza.

Las regiones de rechazo se establecen mediante un *nivel de significancia* (α). La hipótesis se rechaza si $p - \text{valor} < \alpha$.

A.2.4. Tratamiento de errores

El tratamiento de errores es necesario debido a que se puede no partir de la hipótesis correcta o no se está realizando en las condiciones más ideales.

La *potencia de prueba* (β) indica la capacidad para funcionar de la forma planteada con una probabilidad $1 - \beta$. Por ejemplo, si obtenemos una β igual a 0.05, significa que hay un 5% de probabilidad de que la prueba haya fallado, es decir, un 95% de probabilidad de que la prueba realizada haya sido exitosa.

Existen errores en las hipótesis en el diseño de un experimento:

- **Errores Tipo I:** Se considera falsa a la hipótesis nula cuando es verdadera. La probabilidad de que este error ocurra tiene el mismo valor que el nivel de significancia (α). Por ejemplo un error tipo I del 15% implica un α igual a 0.15. Este error es el denominado *falso positivo*.
- **Errores Tipo II:** Se considera verdadera a la hipótesis nula cuando es falsa. La probabilidad de que este error ocurra tiene el mismo valor que la potencia de prueba (β). Este error es el denominado *falso negativo*.

A.2.5. FDR y q-valores

Al realizar numerosas pruebas de hipótesis que pueden dar lugar a errores de tipo I, puede dar lugar a la distorsión de los resultados finales. Para corregir estos errores se recurre al cálculo de q-valores.

Un q-valor inferior a un determinado valor, mide la FDR a tomar si se acepta la prueba efectuada para un p-valor igual o inferior al utilizado. El q-valor indica el máximo error tipo I en un p-valor determinado.

A.2.6. Calidad del experimento

Es necesaria la utilización de indicadores que nos aporten información sobre la fiabilidad del experimento y sus resultados a la hora de verificar los resultados positivos y negativos. Los indicadores son:

- **Sensibilidad (SEN, Sensibility)**
Los verdaderos positivos entre el total de positivos. Es la suma de los verdaderos positivos y de los falsos negativos (FN). Se define a la sensibilidad como la probabilidad de que las prueba muestre un resultado negativo al no cumplirse la hipótesis de partida.
- **Especificidad (SPE, Specifity)**
Los verdaderos negativos entre el total de los sanos (suma de verdaderos negativos y falsos positivos). Se define como la probabilidad de que la prueba muestre

un resultado negativo si la hipótesis de partida al no cumplirse la hipótesis de partida.

■ **Exactitud (ACC, Accuracy)**

Las muestras verdaderas sobre el total (muestras positivas (P) y negativas (N)). Indica la cercanía del valor real al obtenido en la prueba.

$$SEN = \frac{VP}{VP + FN} \quad SPE = \frac{VN}{VN + FP} \quad ACC = \frac{VP + VN}{N + P}$$

Figura A.2: Indicadores de calidad.

Cuando se quiere alcanzar un buen análisis, se busca tanto la sensibilidad como la especificidad con valores próximos 1. La relación entre los 2 valores se establece a través de las curvas ROC. Estas curvas sitúan la especificidad en el eje horizontal y la sensibilidad en el eje vertical.

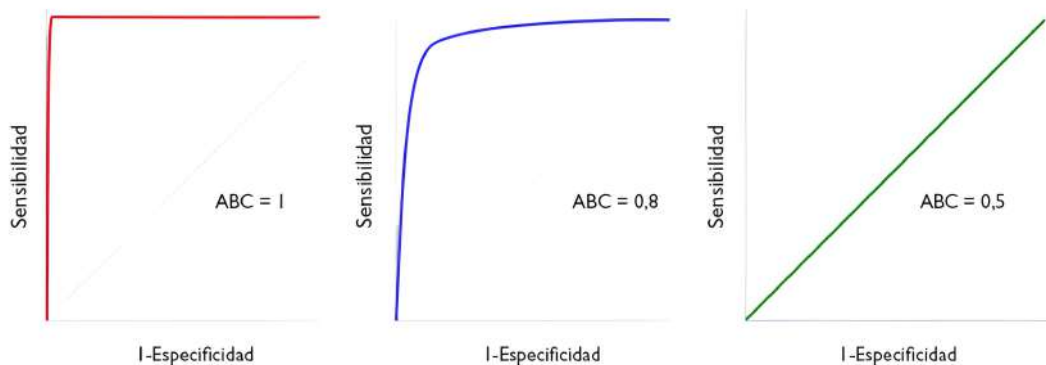


Figura A.3: Ejemplos de curvas ROC de calidades.

A.3. Modelos de análisis de la varianza

El Análisis de la Varianza (ANOVA, *ANalysis Of VAriance*) es una herramienta utilizada para tratar una serie de variables pertenecientes a distintos grupos. Su objetivo es evaluar en qué medida un sólo marcador estadístico permite discernir entre las distintas poblaciones de estudio.

El ANOVA se utiliza tanto en el diseño de experimentos como en el contraste de hipótesis. Estos experimentos pretenden evaluar el efecto de la modificación de una variable en la variabilidad de una respuesta. Estos análisis están condicionados tanto

por la naturaleza del experimento como por las condiciones de diseño deseadas. Para tratar el análisis ANOVA, hay que conocer:

- **Unidad experimental:** Objeto o magnitud tomador de medidas.
- **Factor:** Variable "independiente" que se puede modificar a voluntad por el autor del experimento.
- **Nivel:** Variable que indica la intensidad de un factor.
- **Tratamiento:** Combinación concreta de los mismos niveles de un factor.
- **Respuesta:** Variable del experimento devuelta y medida.

El análisis ANOVA utiliza una distribución *f de Fischer*, evaluando la correlación entre la variabilidad dentro de grupos y la variabilidad entre grupos. Es necesario que entre cada población, las observaciones de la variable de estudio tengan una distribución normal con la misma varianza (σ^2).

Para representar gráficamente un análisis ANOVA se utilizan los *diagramas de cajas y bigotes* (Figura ??). Estos diagramas están compuestos por:

- **Línea central de mediana:** Línea situada en el valor central de las medidas ordenadas de menor a mayor. Sitúa a cada lado la mitad de las medidas,
- **Líneas de cuartiles:** Líneas que cierran la caja inferior y superiormente. Representan la ubicación del primer y tercer cuartil, dejando el 25 % y el 75 % de la población del estudio.
- **Extremos:** Líneas que representan los valores máximos y mínimos alcanzados por la variable de estudios. Estas líneas se encuentran unidas a la caja por unas líneas perpendiculares denominadas *bigotes*.
- **Valores atípicos:** Puntos alejados de los extremos. Representan valores fuera de lo esperado que no poseen una incidencia lo suficientemente elevada para ser tomados en cuenta.

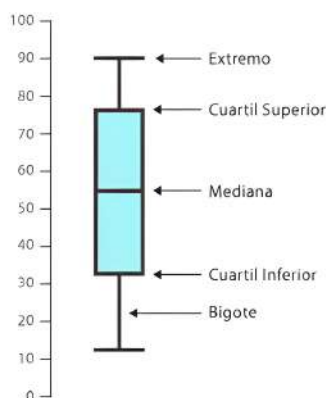


Figura A.4: Partes de un diagrama de cajas y bigotes.

A.4. Métodos no paramétricos.

En ocasiones, se requiere realizar una escala numérica arbitraria que puede indicar qué tan próximos o buenos son los valores medidos respecto a los establecidos como referencia. Para llevar a cabo estos experimentos se utilizan unos métodos denominados *Métodos estadísticos no paramétricos*.

Más allá del análisis de varianzas y medias, estos métodos discriminan entre las poblaciones de estudio. Definen a las distintas poblaciones según cualidades o valores generales y van diferenciando a miembros pertenecientes a unas u otras.

A.4.1. Prueba U de Mann-Whitney.

Es un ensayo no paramétrico aplicado para diferenciar dos muestras independientes. Se diseñó por el químico y estadístico estadounidense Frank Wilcoxon cerca del año 1945, aplicándose inicialmente solo a muestras de tamaño igual o similar. Henry B. Mann y Donald R. Whitney lo extienden a muestras arbitrarias.

La prueba *U de Mann-Whitney* se considera la versión no paramétrica del *test T de Student* para discernir en qué grado dos distribuciones son homogéneas. Son condiciones necesarias que:

- Las observaciones de ambas distribuciones sean independientes.
- Las variables de trabajo sean continuas u ordinales.
- La distribución de ambos grupos tienen el mismo comportamiento ante la hipótesis nula (H_0).
- Si se encuentra en el caso de la hipótesis alternativa (H_1), los valores de una distribución tienen que mostrar una tendencia a ser superiores a los de la otra, es decir, ser distribuciones diferentes.

Hay que destacar el estadístico U , proveniente de establecer el mínimo entre los valores U_1 y U_2 . En las ecuaciones siguientes, n_1 y n_2 son el número de observaciones de cada población, respectivamente. R_1 y R_2 son los correspondientes rangos de cada observación.

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \sum R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum R_2$$

Si hay un número mayor a 20 observaciones en cada muestra, se consideran suficientemente grandes y se aproximan por una distribución normal con media y desviación estándar:

$$\bar{U} = \frac{n_1 n_2}{2}$$

$$\sigma_U = \sqrt{\frac{n_1 n_1 (n_1 + n_2 + 1)}{12}}$$

La distribución normal se define por:

$$z = \frac{U - \bar{U}}{\sigma_U}$$

Ranksum en Matlab

Ranksum es una función en Matlab destinada a prueba de suma de rangos de Wilcoxon.

$p = \text{ranksum}(x, y)$ devuelve el p-valor de una prueba de suma de rangos de Wilcoxon de dos muestras. Esta función prueba la hipótesis nula de que los datos x e y son muestras de distribuciones continuas con medianas iguales, contra la alternativa de que no lo son. Se asume que las dos muestras son independientes y pueden tener diferentes longitudes. Es equivalente a la Prueba U de Mann-Whitney.

En nuestro caso de estudio, un ejemplo se define de la siguiente manera:

$$\text{ranksum}(\text{raw_bsl}(\text{group} == \text{gr}(1), i), \text{raw_bsl}(\text{group} == \text{gr}(2), i))$$

De esta forma obtenemos el p-valor de ambas poblaciones.

A.5. Correlación lineal de Pearson.

El coeficiente de correlación de Pearson es un índice para medir el grado de covariación entre distintas variables relacionadas linealmente (x e y).

Este índice se ejecuta de forma fácil al igual que su interpretación. El coeficiente de Pearson varía en valor absoluto entre cero y la unidad. Si no se consideran estos valores absolutos, efectivamente oscila entre -1 y +1:

- +1: Relación *perfecta positiva*. Cuando aumenta x , aumenta y y cuándo disminuye x , también lo hace y .
- -1: Relación *perfecta negativa*. Cuando aumenta x , disminuye y y cuándo disminuye x , aumenta y .

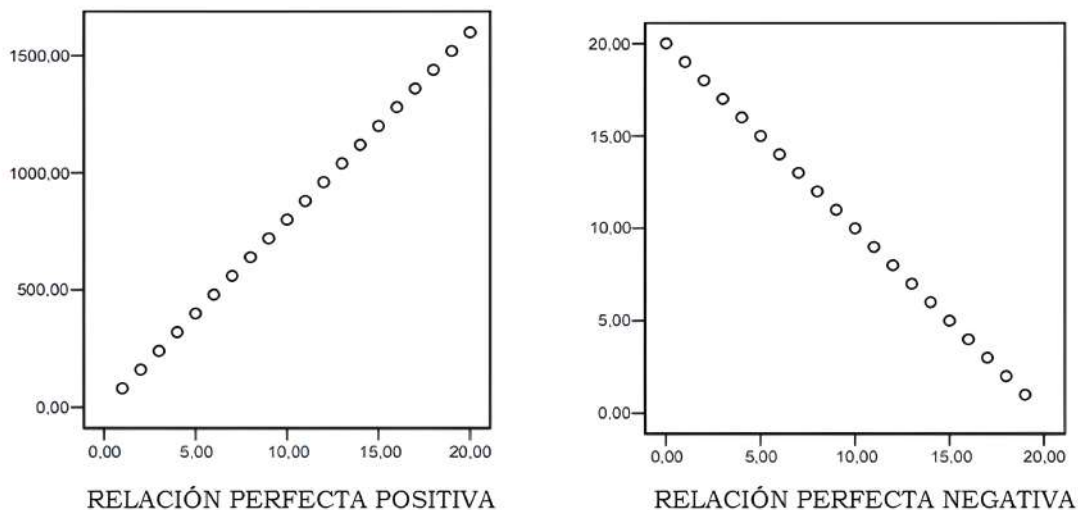


Figura A.5: Relación perfecta positiva y perfecta negativa.

El coeficiente de correlación de Pearson se define por la siguiente expresión:

$$r_{xy} = \frac{\sum Z_x Z_y}{N}$$

Como se puede observar, este coeficiente hace referencia a la media de los productos cruzados de las puntuaciones estandarizadas de X e Y. Refleja la medida en que los individuos se separan de la media.

- Si para un individuo tomamos dos medidas en x e y , la correlación entre estas dos variables será perfecta positiva cuando este individuo tenga la misma superioridad o inferioridad en cada una de ellas, es decir cuando $Z_x = Z_y$.
- Si para un individuo tomamos dos medidas en x e y , la correlación entre estas dos variables será perfecta positiva cuando este individuo tenga la misma superioridad o inferioridad en cada una de ellas, pero en signo contrario, es decir cuando $Z_x = -Z_y$.
- Si para un individuo tomamos dos medidas en x e y , la correlación entre estas dos variables será nula cuando con un valor de Z_x se obtenga cualquier valor de Z_y .

A.5.1. Interpretación del coeficiente de correlación.

Cuanto más cerca de 1, mayor será la correlación y menor cuanto más cerca de cero. Si el coeficiente se encuentra en un valor intermedio, que sea bajo o alto depende de la naturaleza de la investigación.

Hay que tener en cuenta el tamaño de la muestra para considerar característico el coeficiente de correlación.

Corrcoef en Matlab

`corrcoef(A, B)` devuelve coeficientes entre dos variables aleatorias A y B . Como se sabe, el coeficiente de correlación de dos variables aleatorias es una medida de su dependencia lineal.

En nuestro caso de estudio, un ejemplo se define de la siguiente manera:

$$\text{corrcoef}(asyn, abeta)$$

De esta forma obtenemos el coeficiente de correlación entre ambas variables de una población.

Apéndice B

Comparaciones múltiples

B.1. Introducción

En este apéndice, se explican técnicas para analizar en el post-procesado los datos de un experimento tras la realización del Análisis de la Varianza (anova).

Si este Análisis de la Varianza confirma diferencias significativas, se procede a la investigación de qué medias son distintas. Estas técnicas de investigación se denominan *contrastos para comparaciones múltiples*, ya que compara entre sí medias de grupos de ellas.

B.2. Comparaciones múltiples de medias.

El objetivo de estudiar el comportamiento de un factor, mediante un análisis de varianza, es saber si esos grupos difieren significativamente entre sí. Una vez aceptada la existencia de diferencias, hay que conocer cuáles son los grupos diferentes entre sí.

Para ello, primero hay que presentar el modelo de efectos fijos (Ecuación B.1)

$$y_{ij} = \mu_i + u_{ij} \quad (\text{B.1})$$

B.2.1. Comparaciones gráficas de medias

Se considera un modelo equilibrado con n observaciones por nivel y con $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_i$ las medias muestrales de los diferentes grupos. Bajo la hipótesis del modelo, cualquier promedio \bar{y}_i de un grupo se distribuye alrededor de su media μ_i y con una desviación estándar $\frac{\sigma}{\sqrt{n}}$. Si se aceptan las medias de los distintos niveles de un factor como idénticas, las medias se comportan como un conjunto de I observaciones con una distribución con la misma media μ y desviación estándar $\frac{\sigma}{\sqrt{n}}$.

Se pueden distinguir dos casos:

1. Efectos de los tratamientos son iguales entre sí: tienen una muestra aleatoria de la misma distribución.
2. Efectos de los tratamientos no son iguales: no tienen una muestra aleatoria con la misma distribución.

B.2.2. Comparaciones basadas en la distribución t

Las pruebas estadísticas para comparaciones múltiples más frecuentemente utilizadas se basan en la distribución t de Student. Se supone que interesa comparar por parejas los efectos de I grupos. Nos interesa contrastar cualquier hipótesis de la forma:

$$\begin{aligned} H_0 &: \mu_i = \mu_j \\ H_1 &: \mu_i \neq \mu_j \end{aligned}$$

La técnica mas antigua y popular para efectuar estas comparaciones múltiples es el procedimiento LSD¹.

Método de Bonferroni.

En este procedimiento se fija un nivel de significación α que se reparte entre cada una de las comparaciones consideradas y se utiliza la desigualdad de Bonferroni:

$$Pr\left(\bigcup_{m=1}^M A_m\right) \leq \sum_{m=1}^M Pr(A_m) \quad (\text{B.2})$$

Se considera que se quiere realizar estimación por intervalos para las $M = \binom{I}{2}$ comparaciones posibles, cada una al nivel de significación $\alpha^* = \frac{\alpha}{M}$. Esto da origen a M intervalos de confianza con las posibles diferencias $\mu_i - \mu_j$ con probabilidad $1 - \alpha^*$. Si se llama C_m al intervalo m -enésimo:

$$Pr[\mu_{1m} - \mu_{2m} \in C_m] = 1 - \alpha^* \quad m=1,2,\dots,M$$

Siendo μ_{1m} y μ_{2m} la primera y segunda media de la correspondiente comparación. Se supone:

$$1 \leq 1m < 2m \leq I$$

Si se aplica la desigualdad de Bonferroni, se tiene:

$$Pr\left(\bigcap_{m=1}^M C_m\right) = 1 - Pr\left(\bigcap_{m=1}^M \overline{C_m}\right) \geq 1 - \sum_{m=1}^M Pr(\overline{C_m}) = 1 - \sum_{m=1}^M \alpha^* \quad (\text{B.3})$$

Se observa a $\overline{C_m}$ como al complementario de C_m .

Teniendo en cuenta este resultado y si se quiere garantizar un nivel de significancia α , basta tomar:

$$\alpha^* = \frac{\alpha}{M}$$

Con ello, la probabilidad de que todos los intervalos (C_m) contengan a la correspondiente diferencia de medias será de al menos $1 - \alpha$. Los intervalos finalmente resultan:

¹Least Significant Difference

$$\overline{y_{1m}} - \overline{y_{2m}} \pm t \frac{\alpha}{2M} \sqrt{\widehat{S}_R^2 \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}} \right)} \quad (\text{B.4})$$

Donde $\overline{y_{1m}}, \overline{y_{2m}}$ y n_{1m}, n_{2m} son las medias y los tamaños muestrales correspondientes a la comparación m -ésima.

Denotamos $\theta_m = \mu_{1m} - \mu_{2m}$, $m = 1, 2, \dots, M$, una de las M comparaciones lineales por parejas de medias, para las cuales interesa contrastar:

$$H_0 : \theta_m = 0$$

$$H_1 : \theta_m \neq 0$$

Se rechaza H_0 si:

$$|\theta_m| > B_m,$$

y se acepta en el caso contrario.

Donde:

$$B_m = t \frac{\alpha}{2M} \sqrt{\widehat{S}_R^2 \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}} \right)}. \quad (\text{B.5})$$

Si es un modelo equilibrado, los valores de B_m coinciden y se denominan BSD :

$$BSD = t \frac{\alpha}{2M} \sqrt{\widehat{S}_R^2 \left(\frac{1}{n} + \frac{1}{n} \right)}, \quad (\text{B.6})$$

Siendo n el número de observaciones de cada grupo y $t \frac{\alpha}{2M}$ el valor crítico de la distribución t con el mismo número de grados de libertad que la varianza residual, que deja una probabilidad $\frac{\alpha}{2M}$ a su derecha.

***Multcompare* en Matlab**

`multcompare(stats)` devuelve una matriz c de los resultados de la comparación por pares de una prueba de comparación múltiple utilizando la información contenida en la estructura de estadísticas.

`multcompare` también muestra un gráfico interactivo de las estimaciones y los intervalos de comparación.

La media de cada grupo está representada por un símbolo, y el intervalo está representado por una línea que se extiende desde el símbolo. Dos medios de grupo son significativamente diferentes si sus intervalos son disjuntos; no son significativamente diferentes si sus intervalos se superponen. Si utiliza el ratón para seleccionar

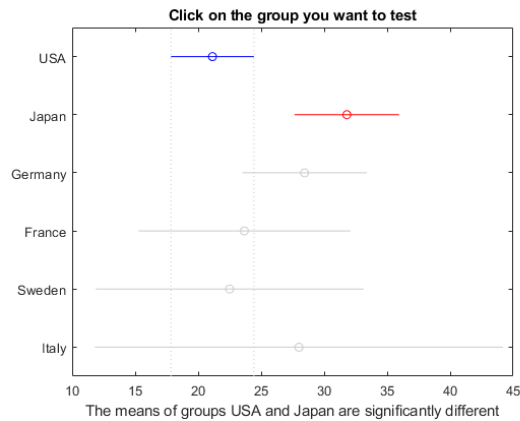


Figura B.1: Ejemplo de gráfico interactivo de estimaciones e intervalos de comparación.

cualquier grupo, el gráfico resaltará todos los otros grupos que son significativamente diferentes, si los hay.

En nuestro caso de estudio, un ejemplo se define de la siguiente manera junto con `anova1`:

```
[p t stats] = anova1(raw_bsl(:, i), group)
[c, m, h, gnames] = multcompare(stats)
title('%s', feature_names{i})
```

Siendo i nuestra característica a comparar entre los grupos y $group$ los tipos de grupos, se obtiene el resultado de si hay diferenciación entre estos en esa característica.

B.2.3. Test de rangos múltiples

Estos contrastes se basan en la distribución del rango estudentizado, cuya definición se realiza según el número de grupos a comparar y de los grados de libertad del estimador de la varianza.

Los métodos resultantes son conservadores, la probabilidad real de rechazar la hipótesis nula cuando es cierta es menor que el nivel de significación α fijado.

Se utiliza para definir al *recorrido* o *rango estudentizado*, k observaciones independientes y_1, y_2, \dots, y_k de una distribución normal de media μ y varianza σ^2 . También se dispone de un estimador S^2 de σ^2 con ν grados de libertad y es independiente de las y_i . Se define a R como el rango de este conjunto de observaciones:

$$R = \max(y_i) - \min(y_i). \quad (\text{B.7})$$

Bajo esto, se define el rango estudentizado:

$$\frac{\max(y_i) - \min(y_i)}{S} = \frac{R}{S}, \quad (\text{B.8})$$

que se denota por $q_{k,\nu}$. La distribución de este estadístico, dependiente de k y ν , ha sido tabulada.

Apéndice C

Empleo de Magerit

C.1. Introducción

En este apéndice, se explica el modo de uso para emplear Magerit-3 en la ejecución de tareas y reducir así, el tiempo de procesado.

C.2. Registro y acceso a Magerit

Antes de todo, lo primero, es realizar la solicitud que se encuentra en:

request.cesvima.upm.es/hpc/upm-learn/

Cuando la solicitud se acepta, se enviará usuario y claves (proyecto, usuario y contraseña).

Para acceder a Magerit-3, debe hacerse de forma remota con la utilización de protocolo SSH, accediendo a *magerit.cesvima.upm.es*.

C.3. SLURM

Para ejecutar tareas que se desean realizar, es necesario lanzarlas a Magerit a través de un fichero *batch* con las directivas SLURM¹. SLURM es un sistema que permite gestionar las tareas y recursos de altas prestaciones, escalabilidad alta, de código abierto y con posibilidad de errores.

SLURM permite convertir ficheros *batch* de sistemas DOS o Windows en ficheros *batch* para sistemas Unix. Estos ficheros tienen especificados:

- Parámetros de ejecución:
 - Número de núcleos.
 - Memoria por CPU.
 - Tareas

¹Simple Linux Utility for Resource Management

- Aplicaciones o directrices a ejecutar.

SLURM realiza distintas funcionalidades con 3 objetivos:

1. Dar acceso a los usuarios a los distintos nodos según el tiempo estimado para las operaciones.
2. Proporciona entorno de trabajo para controlar y monitorizar el trabajo de cada nodo.
3. Administra los recursos mediante un sistema de tareas en cola.

C.4. Mandatos SLURM

Para gestionar SLURM, se utilizan una serie de mandatos que permiten la interacción entre máquina y usuario que permiten controlar y actualizar los trabajos. Los mandatos más utilizados en este trabajo son:

- **sbatch**. Envía trabajo al sistema y devuelve su identificador.
- **squeue**. Listado de todos los trabajos enviados al sistema (en ejecución o en espera).
- **scancel**. Cancelar el trabajo en cola con el identificador introducido o abortar el trabajo en ejecución.
- **sstat**. Muestra estadísticas del trabajo con el identificador introducido al estar en ejecución.
- **sacct**. Muestra información de ejecución de un trabajo ya realizado.

C.5. Carga de módulos

Para ejecutar los distintos programas, se realiza a través de un gestor de módulos, Lmod. Lmod permite preparar el entorno de ejecución y dependencias del programa elegido. Para cargar módulos, solo es necesario escribir en la terminal:

```
module load <nombre_programa/version>
```

Cargado el módulo del programa, se puede lanzar y correr desde el terminal como en cualquier máquina Linux. Hay otros mandatos importantes relacionados con los módulos:

- **module avail**. Muestra aplicaciones disponibles en el clúster.
- **module list**. Muestra aplicaciones cargadas en la ejecución del trabajo actual.

C.6. Parámetros de ejecución

Los parámetros de ejecución tienen como función principal, establecer la identificación del trabajo a realizar y los recursos que se desean reservar. Para identificarse correctamente por parte de SLURM, cada uno tiene que llevar la palabra `#SBATCH`. Los usados en este trabajo son:

- **-job-name**. Nombre del trabajo con finalidad informativa e identificadora.
- **-partition**. Partición a utilizar.
 - *standard*. 600 cores en 160 horas máximo.
 - *debug*. 40 cores en 1 hora máxima.
- **-nodes**. Número de nodos solicitados.
- **-ntasks**. Número de tareas a realizar de forma que cada proceso pueda ejecutarse de forma paralela en un núcleo.
- **-ntasks-per-node**. Tareas asignadas por nodo físico.
- **-cpus-per-task**. Número de núcleos por tarea.
- **-time**. Tiempo máximo de ejecución del trabajo.
- **-mem-per-cpu**. RAM solicitada por CPU en MB.
- **-mem**. RAM solicitada por nodo en MB.
- **-output**. Nombre del fichero en el que se almacenará la salida estándar del trabajo.
- **-error**. Nombre del fichero en el que se almacenará la salida del error del trabajo.
- **-mail-user**. Dirección de correo electrónico donde enviar las actualizaciones de estado de los trabajos.
- **.mail-type**. Tipo de eventos del trabajo de los que el usuario desea ser informado.
- **-chdir**. Directorio de trabajo.

Ejemplo de carga de parámetros de ejecución

```
#!/bin/bash
## _____ Start job description _____
#SBATCH --partition=standard
#SBATCH --job-name=<Name>
#SBATCH --mem-per-cpu=<memory in MB>
#SBATCH --cpus-per-task=<Core number per task>
#SBATCH --time=160:00:00
#SBATCH --nodes=1
#SBATCH --ntasks=<Parallelizations Number>
```

```
#SBATCH --mail-type=ALL,TIME LIMIT 100
#SBATCH --mail-user=example@domainExample.com
#SBATCH --output=out-%j.log
#SBATCH --error=err-%j.log
#SBATCH --chdir=<path/to/execution/script>
## ----- End job description -----
```

C.7. Directorio SCRATCH

Para la computación de altas prestaciones, es recomendable el uso de un directorio llamado *SCRATCH*. Este directorio almacena temporalmente los residuos producidos durante la ejecución del trabajo. Una vez terminado el trabajo, se elimina. Para la creación y destrucción de la carpeta, se sigue un esquema como el siguiente:

Uso de directorio SCRATCH

```
# Create a temporary directory on scratch
mkdir -p $SCRATCH/$SLURM JOB ID
<execution script>
# Cleanup local work directory
rm -rf $SCRATCH/$SLURM JOB ID
```

C.8. Notificación de estado y redirección de salidas

Es posible añadir una dirección de correo electrónico para recibir notificaciones del estado del trabajo en Magerit.

Se pueden encontrar opciones para la propia ejecución del trabajo:

- *BEGIN*.
- *END*.
- *FAIL*.
- *ALL*.

Otras opciones que se pueden encontrar, son las referentes a los tiempos de procesamiento:

- *TIME_LIMIT*. Se ha consumido el 100% de tiempo asignado.
- *TIME_LIMIT_80*.
- *TIME_LIMIT_50*.

Toda la salida por texto realizada por el código fuente es guardada en el fichero *.log* del campo *output*. También existe un fichero de error en el campo *error*.

Apéndice D

Diccionario de siglas y acrónimos

D.1. Siglas

PDA+, **PDA β +**. Parkinson diagnosticado con depósitos de A β .

PDA-, **PDA β -**. Parkinson diagnosticado sin depósitos de A β .

A β . Beta-amiloide.

ABetaTAU. Ratio A β /tTau.

ACC. Exactitud (Accuracy).

AD. Enfermedad de Alzheimer (*Alzheimer's Disease*).

ADAS-Cog. Alzheimer's Disease Assesment Scale-Cognitive Subscale.

ADL. Activities of Daily Living Scale.

ADNI. Iniciativa de neuroimagen de la enfermedad de Alzheimer (*Alzheimer's Disease Neuroimaging Initiative*).

ANOVA. Análisis de Varianza (*ANalysis Of VAriance*).

AUC. Área bajo la curva (*Area Under the Curve*).

CDR. *Clinical Dementia Rating*.

CeSViMa. Centro de Supercomputación y Visualización de Madrid.

CN. Cognición normal.

CSF Líquido cefalorraquídeo (*Cerebrospinal fluid*).

CV. Validación cruzada (*Cross Validation*)

DAT. Déficit de transporte de dopamina.

de Novo. Diagnóstico de Parkinson temprano sin medicación prevoa contra el mismo.

DPM. Modelo de Progresión de Enfermedad (*Disease Progression Model*).

EMCI. Declive cognitivo leve temprano (*Early Mild Cognitive Impairment*).

ETSIDI. Escuela Técnica Superior de Ingeniería y Diseño Industrial.

FAQ. *Functional Activities Questionary*.

FDR. Tasa de falso descubrimiento (*False Discovery Rate*).

GDS. *Geriatric Depression Scale*.

GLM. Modelo general lineal (*General Linear Model*).

GRACE. *GRrowth models via Alternating Conditional Expectation*.

GUI. Interfaz gráfica de usuario (*Graphic User Interface*).

HC. Sujeto sano de control (*Healthy Control*).

HPC. Computación de altas prestaciones (*High Performance Computing*).

HV. Volumen hipocampal (*Hippocampal Volume*).

HVLT. *Hopckins Verbal Learning Test*.

HVLT-R. *Hopckins Verbal Learning Test-Revised*.

HVLTRDLY. *Hopkins Verbal Learning Test Delayed Recall*.

INMEDIATE. *Hopkins Verbal Learning Test Inmediate*.

JLO. *Benton Judgement of Line Orientation Score*.

LDA. Análisis Linear Discriminante (*Linear Discriminant Analysis*).

LMCI. Declive cognitivo leve tardío (*Late Mild Cognitive Impairment*).

LME. Modelo de efectos lineales mixtos (*Lineal Mixed Effects*).

LNS. *Letter Number Sequencing Score*.

- MAE.** Error Medio Absoluto (*Medium Absolute Error*).
- MATLAB.** *MATrix LABoratory*.
- MCI.** Declive cognitivo leve (*Mild Cognitive Impairment*).
- MEANCAU.** Absorción media en caudado.
- MEANPUT.** Absorción media en putamen.
- MMSE.** Mini Prueba del Estado Mental (*Mini Mental State Examination*).
- MoCA.** Evaluación Cognitiva de Montreal (*Montreal Cognitive Assessment*).
- MRI.** Imagen de Resonancia Magnética (*Magnetic Resonance Image*).
- mRMR.** Mínima redundancia y máxima relevancia (*minimal-Redundancy-Maximal-Relevance*).
- NaN.** Dato no numérico (*Not a Number*).
- NAV.** Volumen normalizado de la amígdala (*Normalized Amygdala Volume*).
- NHV.** Volumen hipocampal normalizado (*Normalized Hippocampal Volume*).
- NM.** Medidas neuropsicológicas (*Neuropsychological measurements*).
- PD.** Enfermedad de Parkinson (*Parkinson's Disease*).
- PD-MCI+.** Diagnóstico de Parkinson con deterioro cognitivo leve.
- PD-MCI-.** Diagnóstico de Parkinson sin deterioro cognitivo leve.
- PDD.** Diagnóstico de Demencia de Parkinson (*Parkinson's Disease with Dementia*).
- PET.** Tomografía de Emisión de positrones (*Positron Emission Tomography*).
- PPMI.** Iniciativa de Marcadores de Progresión del Parkinson (*Parkinson's Progression Markers Initiative*).
- pTau.** tau fosforilada (*phosphorylated tau*).
- RAVLT.** *Rey Auditory Verbal Learning Test*.
- RBDSQ.** *Rapid Eye Movement Sleep Behaviour Disorder Screening Questionnaire*.

REM Movimiento Rápido del Ojo (*Rapid Eye Movement*).

RF Radiofrecuencia.

ROC Característica Operativa del Receptor (*Receiver Operating Characteristic*).

ROI. Región de Interés (*Region Of Interest*).

RSS. Suma de cuadrados residual (*Residual Sum of Squares*).

SCOPA. *Scales for Outcomes in Parkinson's Disease*.

SDM. *Symbol-Digit Modalities Score*.

SDMT. *Symbol-Digit Modalities Test*.

SEN. Sensibilidad (*Sensibility*).

SFT. *Semantic Fluency Total Score*.

sHC. Sujeto de sano de control estable (*Stable Healthy Control*).

SLURM. *Simple Linux Utility for Resource Management*.

sMCI. Sujeto con declive cognitivo leve estable (*Stable Mild Cognitive Impairment*).

SPE. Especificidad (*Specifity*).

SPECT. Tomografía Computarizada de Emisión Monofotónica (*Single Photon Emission Computed Tomography*).

ST-LME. Modelo de efectos lineales mixtos espacio-temporal (*Spatiotemporal Lineal Mixed Effects*).

SUV. Valor de absorción estándar (*Standardized Uptake Value*).

SUVR. Valor de absorción estándar de ratio compuesto (*Standardized Uptake Value*).

tTau. tau total (*total tau*).

UPDRS *Movement Disorder Society-United Parkinson's Disease Rating Scale Total Score*.

UPM. Universidad Politécnica de Madrid.

UPSIT. *University of Pennsylvania Smell Identification Test.*

V-LME. Modelo de efectos lineales mixtos en vértices (*Vertex Lineal Mixed Effects*).

Bibliografía

- [1] Miguel Lerín Alonso. Estudio longitudinal mediante biomarcadores y neuroimagen para desarrollo de modelos predictivos de conversión a declive cognitivo en pacientes parkinson. 2021.
- [2] Zacharias Anastasiou Daniel Weintraub Jonathan M Schott Anette Schrag, Uzma Faisal Siddiqui. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed parkinson's disease: a cohort study. *The Lancet Neurology*, 16(2):66–75, 2016.
- [3] MoCA. MOCA Montreal Cognitive Assessment. <https://www.mocatest.org/>.
- [4] CeSViMa. Guía del usuario hpc. documentación de hpc. <https://docs.cesvima.upm.es/docs/magerit/es/latest/>. online. 2020.
- [5] Brenna Cholerton, Catherine O Johnson, Brian Fish, Joseph F Quinn, Kathryn A Chung, Amie L Peterson-Hiller, Liana S Rosenthal, Ted M Dawson, Marilyn S Albert, Shu-Ching Hu, et al. Sex differences in progression to mild cognitive impairment and dementia in parkinson's disease. *Parkinsonism & related disorders*, 50:29–36, 2018.
- [6] David A. Bennettb Kaj Blennowc Maria C. Carrillod Billy Dunne Samantha Budd Haeberleinf David M. Holtzmang William Jagusth Frank Jesseni Jason Karlawishj Enchi Liuk Jose Luis Molinuevol Thomas Montinem Creighton Phelps Katherine P. Rankino Christopher C. Rowep Philip Scheltensq Eric Siemersr Heather M. Snyderd Reisa Sperlings Clifford R. Jack, Jr. NIA-AA research framework: Toward a biological definition of alzheimer's disease. *2018 National Institute on Aging—Alzheimer's Association (NIA-AA) Research Framework*, pages 535–562, 2018.
- [7] Glenda M. Halliday Gert J. Geurtsen Clive Ballard K. Ray Chaudhuri Dag Aarsland, Lucia Batzu and Daniel Weintraub. Parkinson disease-associated cognitive impairment. *NATURE REVIEWS*, pages 1–21, 2021.
- [8] Janel Fedler PhD Christopher S. Coffey PhD Chelsea Caspell-Garcia MS Ju Hee Kang PhD Tanya Simuni MD Tatiana Foroud PhD Arthur W. Toga et al David J. Irwin, MD. Evolution of alzheimer's disease cerebrospinal fluid biomarkers in early parkinson's disease. *Annals of Neurology*.
- [9] Universidad de Barcelona. Inferencia estadística. http://www.ub.edu/aplica_infor/spss/cap4-1.htm.
- [10] Sociedad Española de Bioestadística. La estadística en ciencias de la salud. una mirada desde la investigación. *Universidad de Extremadura*.

- [11] Coeficiente de correlación de Pearsonl. [https://es.wikipedia.org/wiki/coeficiente_de_correlaci %c3](https://es.wikipedia.org/wiki/coeficiente_de_correlaci%C3)
- [12] Contraste de hipótesis. [http://wpd.ugr.es/ bioestad/guia-de-r/practica-6/](http://wpd.ugr.es/bioestad/guia-de-r/practica-6/). 2021.
- [13] Ministerio de Sanidad. Patrones de mortalidad en españa, 2018. *Ministerio de Sanidad*, 2022.
- [14] El Club del Autodidacta. Latex. <http://elclubdelautodidacta.es/wp/indice-latex/>.
- [15] Economipedia. Inferencia estadística. <https://economipedia.com/definiciones/inferencia-estadistica.html>.
- [16] Diego Cecchin Anna Chiara Frigo Jinhee Kim Luca Weis-Antonio P Strafella Eleonora Fiorenzato, Roberta Biundo and Angelo Antonini. Brain amyloid contribution to cognitive dysfunction in early-stage parkinson’s disease: the ppmi dataset. *Journal of Alzheimer’s Disease*, 66(1):229–237, 2018.
- [17] Florentino Prado Esteban and Angelica Muñoz Pascual. De parkinson.
- [18] Filezilla. https://filezilla-project.org/client_features.php.
- [19] Filezilla. <https://www.xataka.com/basics/ftp-que-como-funciona>.
- [20] FileZilla. Filezilla. the free ftp solution. <https://filezilla-project.org/>. 2022.
- [21] Inés González. Parkinson: El número de afectados se duplicará en 20 años y se triplicará en 2050. *Sesenta y más*, (339):52– 57, 2018.
- [22] Pedro Gil Gregorio. Neurodegeneración. alzheimer, parkinson y ela. 2018.
- [23] Colin Harrison. Xming x server for windows. oficial website. <http://www.straightrunning.com/xmingnotes/>. 2020.
- [24] John Hobson. The montreal cognitive assessment. *Occupational Medicine*, 9(65):764–765, 2015.
- [25] Margaret M. Hoehn and Melvin D. Yahr. Parkinsonism onset, progression, and mortality. *Neurology*, 17(5).
- [26] Douglas N Greve Bruce Fischl Mert R Sabuncu Alzheimer’s Disease Neuroimaging Initiative et al. Jorge L Bernal-Rusiel, Martin Reuter. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage*, 81:358–370, 2013.
- [27] Martin Reuter Bruce Fischl Jorge L. Bernal-Rusiel, Douglas N. Greve and Mert R. Sabuncu. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage*, 66:249–260, 2013.
- [28] Leslie Lamport. Latex. a document preparation system. <https://www.latexp.org/>. 1986.
- [29] Magerit. <https://docs.cesvima.upm.es/magerit/jobs/>.

- [30] Magerit. <https://www.cesvima.upm.es/infrastructure/magerit>.
- [31] Mathworks. Matlab and simulink. <https://www.mathworks.com/products/matlab.html>. 2022.
- [32] Matlab. <https://es.wikipedia.org/wiki/matlab>.
- [33] MD; David P. Salmon PhD; Dorene M. Rentz et al. Michael C. Donohue, PhD; Reisa A. Sperling. The preclinical alzheimer cognitive composite. measuring amyloid-related decline. *American Medical Association*.
- [34] M´elanie Le Goff Ronald G Thomas Rema Raman Anthony C Gamst Laurel A Beckett Clifford R Jack Jr Michael W Weiner Jean-Fran,cois Dartigues et al Michael C Donohue, Helene Jacqmin-Gadda. Estimating long-term multivariate progression from short-term data. *Alzheimer’s and Dementia*, 10:400–410, 2014.
- [35] Han-Kyeol Kim³ & Chul Hyoung Lyoo Min Seok Baek, Myung Jun Lee. Temporal trajectory of biofluid markers in parkinson’s disease. *Nature*.
- [36] Jesus Montero and Carmen Minuesa. Estadística básica para las ciencias de la salud. *Universidad de Extremadura*.
- [37] Juan Carlos Correa Morales and Juan Carlos Salazar Uribe. Introducción a los modelos mixtos. *Escuela de Estadística*, 2016.
- [38] University of Chicago. User guide matalb. <https://rcc.uchicago.edu/docs/software/environments/matlab/>. 2014.
- [39] OpenSSH. <https://www.openssh.com/>.
- [40] Celia Painous and Maria J Marti. Cognitive impairment in parkinson’s disease: what we know so far. *Journal of Parkinsonism & Restless Legs Syndrome*, 10:7, 2020.
- [41] Personal. Coeficiente de correlación lineal de pearson. <https://personal.us.es/vararey/adatos2/correlacion.pdf>.
- [42] Carlos Platero. Categorical predictive and disease progression modeling in the early stage of alzheimer’s disease. *Preprint submitted to Journal of Neuroscience Methods*, 66(1):1–38, 2021.
- [43] Parkinson’s progression markers initiative PPMI. <https://www.ppmiinfo.org/about-ppmi/>. 2022.
- [44] Putty. <https://blog.baehost.com/uso-de-putty-y-comandos-frecuentes-en-ssh/>.
- [45] Putty. <https://elblogdeladministrador.com/2017/06/26/mapear-xwindows-con-putty-xming/>.
- [46] Putty. <https://microtecnologias.wordpress.com/2008/04/14/xmingputty-administracion-grafica-y-remota-de-un-servidor-linux/>.
- [47] PuTTY. Download putty - a free ssh and telnet client for windows. <https://www.putty.org/>. 2020.
- [48] Marina Ruiz Pérez. Estudio longitudinal del deterioro cognitivo en la enfermedad de parkinson. 2021.

- [49] Pere Rebas. Conceptos basicos del análisis de supervivencia. *Cirugía Española*, 78(4).
- [50] Pere Rebas. nova analisis de varianza para comparar multiples medias.
- [51] Michele Romoli and Paolo Calabresi. Como nos movemos. un viaje al sistema cerebral que controla el movimiento. 2019.
- [52] Francesc Vallderiola Serra and C. Gaig Ventura. Actualización en la enfermedad de parkinson. *Neurologia Suplementos*, 2(1):10–18, 2006.
- [53] Nicola Smith, Owen A Williams, Lucia Ricciardi, Francesca Morgante, Thomas R Barrick, Mark Edwards, and Christian Lambert. Predicting future cognitive impairment in de novo parkinson’s disease using clinical data and structural mri. *medRxiv*, 2021.
- [54] Chunyan Tang, Xiaoyan Zhao, Wei Wu, Weijia Zhong, and Xiaojia Wu. An individualized prediction of time to cognitive impairment in parkinson’s disease: A combined multi-predictor study. *Neuroscience Letters*, 762:136149, 2021.
- [55] MathWorks Support Team. Why am i unable to use parpool with the local scheduler or validate my local configuration of parallel computing toolbox? 2012.
- [56] Kostas M Tsiouris, Spiros Konitsiotis, Dimitrios D Koutsouris, and Dimitrios I Fotiadis. A risk stratification model for early cognitive impairment after diagnosis of parkinson’s disease. In *Mediterranean Conference on Medical and Biological Engineering and Computing*, pages 653–660. Springer, 2019.
- [57] Solicitud HPC upm-learn – CeSViMa. <https://request.cesvima.upm.es/hpc/upmlearn/>.
- [58] Wikipedia. Curva roc.
- [59] Robert Beaver William Mendenhall and Barbara Beaver. Introduccion a la probabilidad y estadistica.
- [60] Xming. <https://sourceforge.net/projects/xming/>.
- [61] M Elena Toribio-Díaz y Francisco J Carod-Artal. Subtipos de deterioro cognitivo leve en la enfermedad de parkinson y factores predictores de conversión a demencia. *Rev Neurol*, 61:14–24, 2015.
- [62] Millán Arroyo Menendez y Lucila Finkel Morgenstern. Dependencia e impacto social de la enfermedad de parkinson. *Revista española de discapacidad*, 2:25–49, 2013.
- [63] Carlos Platero y M. Carmen Tobar. Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to alzheimer’s disease. *Journal of Neuroscience Methods*, January(341):108698, 2020.