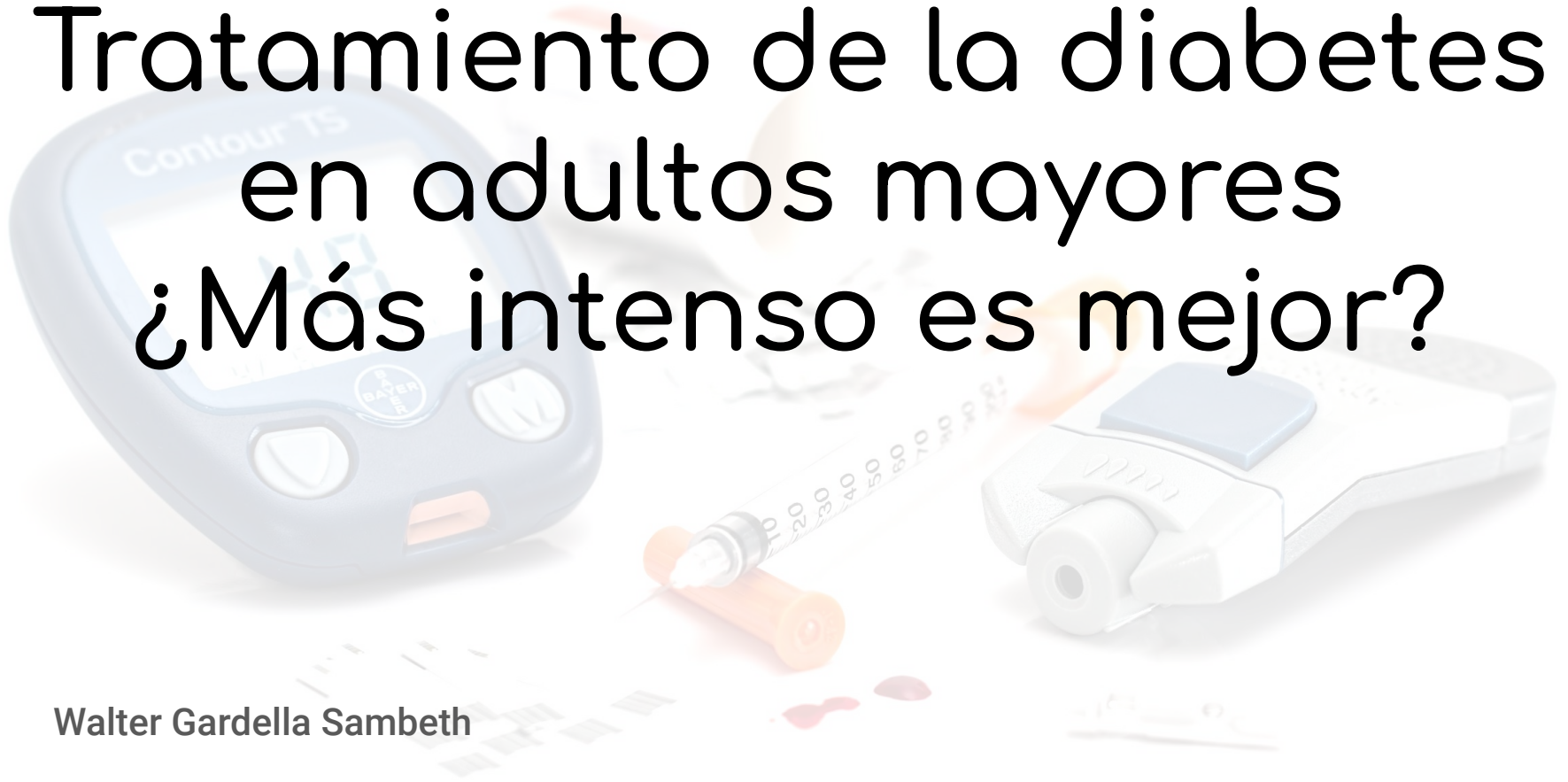


Tratamiento de la diabetes en adultos mayores ¿Más intenso es mejor?

Walter Gardella Sambeth



Resumen

- Los consensos internacionales acerca del tratamiento de la diabetes abogan por enfoques intensificados de terapia para prolongar la sobrevida del paciente.
- Sin embargo, en un paciente añoso, el riesgo a corto plazo inherente al tratamiento puede que no compense los posibles beneficios a largo plazo por la poca esperanza de vida.
- Se propone modelar la tasa de mortalidad a 5 y 10 años de pacientes diabéticos añosos con y sin comorbilidades a fin de evaluar la relación costo/beneficio de dichos tratamientos.

Preguntas

- ¿Hay una diferencia en la tasa de mortalidad de un adulto mayor con diabetes con y sin otras patologías?
- ¿Qué variables afectan significativamente esa tasa de mortalidad?

Datos empleados

Se utilizó un dataset obtenido de la página [mendeley.com](https://www.mendeley.com).

Contiene 275190 filas por 70 columnas describiendo información recopilada de veteranos de las Fuerzas Armadas norteamericanas mayores de 65 años.

Registra parámetros demográficos, medidas antropométricas y datos de laboratorio respecto de su salud al momento del estudio, y se registró si habían fallecido a los 5 y 10 años luego de la entrevista original.

Análisis exploratorio

Clases de variables

14 variables numéricas

56 variables categóricas:

- 53 variables binarias
- 3 variables con valores múltiples

Análisis exploratorio

Consideraciones

- Los pacientes tenían 65 años o más de edad al 1/1/2006.
- Ingresados al programa de atención de salud a Veteranos de las Fuerzas Armadas.
- Todos tenían diabetes diagnosticada clínicamente al momento de su inclusión en el ensayo.
- Al menos una visita al médico con controles registrados para tres parámetros típicos de la diabetes (Tensión arterial, IMC y hemoglobina A1C) en los 24 meses anteriores al estudio.
- El sexo sólo se considera binario Hombre/Mujer.
- El estado civil sólo se considera Casado, Soltero o Viudo.
- Se definieron ocho grupos de prioridad para la inclusión en el estudio teniendo en cuenta niveles de discapacidad física y dificultades socioeconómicas.

Análisis exploratorio

PRESENCIA DE NULOS EN EL DATASET

- La columna MICROALB presenta un 72% de valores nulos. Esto se debe a que es un parámetro que únicamente tiene valores medibles en pacientes con daño renal.
- De forma similar, la columna SERUMALB tiene un 30% de nulos porque no es un valor que se mida de forma habitual, sino que el médico debe tener motivos para solicitar el control.
- El resto de las columnas tiene una cantidad despreciable de nulos

Hallazgos en el dataset

- El dataset tiene un 1,69% de registros duplicados. Dada la baja proporción, fueron eliminados antes de continuar.
- Las mujeres representan una muy pequeña minoría de los registros.
- Más del 75% de los registros corresponden a personas de raza blanca.
- La gran mayoría de los pacientes es hipertenso y recibe medicación para esa enfermedad.
- Las variables para obesidad y bajo peso se desestimaron porque sus valores eran incoherentes.
- Casi la cuarta parte de los pacientes falleció dentro de los 5 años del estudio.
- Más de la mitad de los pacientes falleció dentro de los 10 años del estudio.
- Las variables numéricas tienen una dispersión muy elevada, lo que hace complejo analizar el efecto de cada una en la mortalidad total.
- Algunas variables presentan muchos outliers, pero se mantienen por su significación médica.

Conversión de variables

Para simplificar el modelo, se realizan las siguientes conversiones en las variables:

- Todas las variables binarias se convierten a valores distintos de 0 (1=Falso, 2=Verdadero)
- Las variables numéricas se agrupan en función de su significación médica en valores enteros (desde 1 a 2, 3, 4 o 5 según la variable)
- Todos los valores nulos se convierten a 0

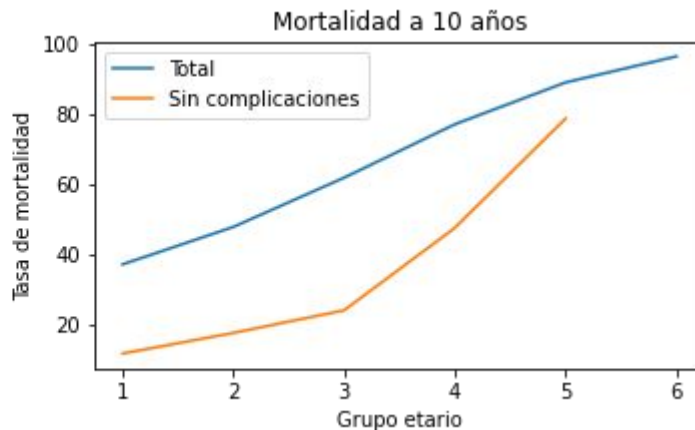
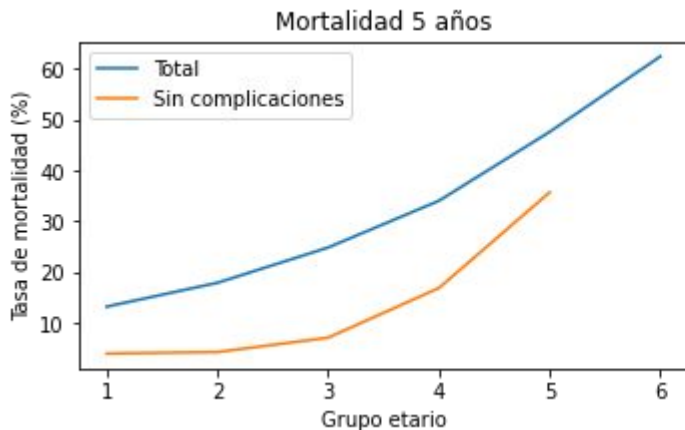
Correlaciones entre variables

El análisis de las correlaciones entre las variables convertidas muestra 9 grupos que se correlacionan con fuerza:

- Ocho de los grupos se explican de acuerdo a la significación médica de las variables.
- Uno de los grupos (Colesterol HDL / Triglicéridos) no parece tener una explicación médica inmediata.

Cálculo de la tasa de mortalidad

Se calculó la tasa de mortalidad para cada grupo etario en general y únicamente para los pacientes sin complicaciones:



Se puede apreciar que en todos los grupos etarios la tasa de mortalidad fue menor para los pacientes sin comorbilidades.

Evaluación de outliers

Se evaluaron por separado la proporción de outliers en los conjuntos de datos para la predicción a 5 y 10 años utilizando entre los métodos Isolation Forest, Minimum Covariance Determinant, Local Outlier Factor y One-Class SVM.

Se encontró que el mejor método para detectar los outliers en la predicción a 5 años fue One-Class SVM y para la predicción a 10 años fue Isolation Forest.

La proporción de outliers fue baja (<3.5%) pero se los eliminó de todas formas. El impacto en el R^2 de las correlaciones no fue significativo.

Análisis de Componentes Principales

Se hizo un análisis de componentes principales para tratar de simplificar el modelo. Simultáneamente se evaluó el efecto de mantener y eliminar los grupos de variables de alta correlación detectados durante el EDA.

	Componentes necesarios para explicar...	
	60% de la varianza	80% de la varianza
Con variables de alta correlación	31	46
Sin variables de alta correlación	29	45

En conclusión, el PCA no logró reducir lo suficiente la complejidad del modelo.

Selección de variables significativas

Se realizó una selección primaria de variables significativas para la regresión lineal de la tasa de mortalidad a 5 y 10 años a partir de las variables convertidas.

Para el proceso se compararon dos métodos:

- El método stepwise con un p-valor para considerar significativa a cada variable de 0.05.
- El método BorutaPy. Este método requiere que la variable objetivo tenga valores discretos por lo que se escalaron las tasas de mortalidad a fin de retener 5 cifras significativas.

Insights obtenidos

- La presencia de comorbilidades implica una mayor tasa de mortalidad.
- La conversión de las variables mejora ligeramente la correlación: Esto implica que los mecanismos de conversión mantienen el significado médico de cada dato.
- Se observó que cada método de selección de variables parece ser más parsimonioso en una sola de las regresiones, por lo que habrá que definir a posteriori cuál es el conjunto de variables que se usa en cada una.

Aplicación de algoritmos de ML

Para realizar las regresiones se aplicaron los siguientes algoritmos:

- SGDRegressor
- LinearSVR

Ambos tienen la ventaja para nuestro caso de que devuelven los coeficientes de cada variable significativa.

Se trabajó con un training set del 33% del total (Calculado por separado para la regresión a 5 y 10 años y mantenido en todo el proceso).

Optimización de los algoritmos

Se evaluó la performance de los algoritmos elegidos para ajustar las correlaciones con los valores por omisión para las rutinas y luego de optimizar los parámetros por medio de GridSearchCV:

Correlación	Método	R^2	AR^2	MAE	RMSE
5 años	SGDRegressor (básico)	0.951	0.951	0.0143	0.0207
	SGDRegressor (optimizado)	0.951	0.951	0.0148	0.0206
	LinearSVR (básico)	0.944	0.944	0.0155	0.0221
	LinearSVR (optimizado)	0.951	0.951	0.0149	0.0206
10 años	SGDRegressor (básico)	0.993	0.993	0.0107	0.0132
	SGDRegressor (optimizado)	0.994	0.994	0.0105	0.0119
	LinearSVR (básico)	0.988	0.988	0.0133	0.0169
	LinearSVR (optimizado)	0.994	0.994	0.0105	0.0119

Resultados finales

Para los resultados finales se utilizó el algoritmo LinearSVR porque resultó mucho más rápido que SGDRegressor, utilizando el conjunto de variables seleccionado por BorutaPy para la predicción a 5 años y el obtenido por Stepwise para la correlación a 10 años.

Se calculó la relación R entre los coeficientes de la correlación lineal para los pacientes sin comorbilidades y para los que tenían comorbilidades.

También se estudió si cambiaban de signo entre ambas ecuaciones, lo que permitió encontrar tres tipos de casos:

- a) Coeficientes negativos para pacientes con comorbilidades que pasan a ser positivos en pacientes sin comorbilidades, o sea que aumentan mucho la mortalidad (se indicaron como ‡).
- b) Coeficientes que mantienen su signo entre correlaciones.
- c) Coeficientes positivos para pacientes con comorbilidades que pasan a ser negativos (se indicaron como ♥).

Resultados finales - Predicción a 5 años

Variable	Relación R	Indicador
Arritmia cardíaca	43.40	‡
Uso de Biguanidas	6.53	‡
IMC	2.93	‡
Enf. cardíaca congestiva	58.92	
Tabaquismo	7.27	
Triglicéridos en sangre	6.17	
Índice de fragilidad (1)	3.27	
Presión diastólica	2.09	
Edad	0.93	
Prioridad de inclusión	0.39	
Casado/a	1.68	♥

(1): Indica la presencia o ausencia de treinta problemas de salud relacionados con la edad.

Resultados finales - Predicción a 10 años

Variable	Relación R	Indicador
Prob. circulación pulmonar	253.99	‡
Uso de insulina	63.96	‡
Uso de Tiazolidinedionas	48.76	‡
Uso de Biguanidas	40.33	‡
IMC	5.69	‡
Sexo	5.60	‡
Hemoglobina A1C	362.50	
Enfermedad valvular	92.36	

Variable	Relación R	Indicador
Tabaquismo	76.87	
Depresión clínica	59.26	
Prioridad de inclusión	1.72	
Raza	1.65	
Presión diastólica	1.55	
Edad	0.99	
Casado/a	0.39	

Conclusiones

- Se pudo observar una clara diferencia en la tasa de mortalidad, tanto a 5 como a 10 años, entre los diabéticos mayores con complicaciones y aquellos sin ellas.
- Se pudieron diferenciar aquellas variables que explican satisfactoriamente las tasas de mortalidad observadas, y se encontraron diferencias muy claras en el impacto de las mismas en pacientes que tienen comorbilidades y aquellos que no las presentan.
- Este conocimiento permitiría concentrar los esfuerzos de diagnóstico y tratamiento en aquellas comorbilidades que tienen el mayor impacto real sobre la salud del paciente, lo que reduciría la posibilidad de aparición de efectos secundarios sobre un paciente ya frágil por su edad.

A collection of medical supplies for diabetes management, including a blue Contour TS glucose meter, an orange syringe, a white insulin pump, and various test strips and pills, all arranged on a white background with the text 'MUCHAS GRACIAS' overlaid.

MUCHAS GRACIAS