



# Neurology perspectives



## 19840 - Influencia de las técnicas de *undersampling* para la optimización de la clasificación en modelos predictivos: aplicación de caso a la mortalidad por ictus isquémico

Amaya Pascasio, L.<sup>1</sup>; Lea Pereira, M.<sup>2</sup>; Martínez Sánchez, P.<sup>1</sup>; Iglesias Espinosa, M.<sup>1</sup>; García Torrecillas, J.<sup>3</sup>

<sup>1</sup>Servicio de Neurología. Hospital Torrecárdenas; <sup>2</sup>Servicio de Medicina Interna. Complejo Hospitalario de Poniente; <sup>3</sup>Servicio de Urgencias. Hospital Torrecárdenas.

### Resumen

**Objetivos:** Determinar la modificación en la capacidad discriminativa, en sensibilidad y especificidad de un modelo para estimar riesgo de mortalidad por ictus isquémico mediante técnicas de *undersampling*.

**Material y métodos:** Estudio de cohorte histórica conformada por los episodios de hospitalización por ictus isquémico no reperfundido. Fuente: conjunto mínimo básico de datos, periodo 2016-2018 (n = 62 552). Se obtuvo un modelo basal (MB1) del que se estimó su área bajo la curva, sensibilidad (S), especificidad (E), valor predictivo negativo (VPN) y clasificación correcta. Se muestreó aleatoriamente la clase mayoritaria (vivos) y se aplicó el MB1 sobre una base de datos conformada por todos los *exitus* y una selección aleatoria de vivos (proporción 1:3).

**Resultados:** El MB1 obtuvo AUC de 0,731, S: 37,5%, E: 89,4%, VPN: 99,88%, clasificando correctamente el 89,3%. Este modelo se construyó sobre el n total, reportando un 10,6% de mortalidad, siendo varones el 53,5%. MB1:  $\text{Prob}(\text{exitus}) = 1/1 + e^{\text{EXP}(\text{FX})}$ , siendo  $\text{FX} = 0,0001 + 1,073 \cdot \text{edad} + 1,143 \cdot \text{sexo} + 1,192 \cdot \text{cardiopatía isquémica} + 0,719 \cdot \text{hipertensión} + 0,652 \cdot \text{hiperlipemia} + 1,414 \cdot \text{fibrilación auricular} + 2,133 \cdot \text{insuficiencia cardíaca} + 0,755 \cdot \text{ictus territorio posterior}$ . La aplicación del MB1 sobre la base de datos con *undersampling* 1:3 (n = 62.552, 51,4% varones, 25% *exitus*) proporcionó un nuevo modelo con AUC:0,729, S: 58,4%, E: 77,6% y VPN: 96%. Clasificó correctamente al 76,2% de casos.

**Conclusión:** Las técnicas de submuestreo de la clase mayoritaria son ampliamente utilizadas para el manejo de bases desbalanceadas. En el caso del ictus, permitió un discreto incremento de la S a expensas de empeorar la E y porcentaje de clasificación correcta. Se infiere de ello que el *undersampling* modifica seriamente la representación epidemiológica natural del desenlace (*exitus*/vivos) sin alcanzar una tabla de clasificación óptima.