

**Universidad del Bío-Bío**

**Facultad de Ciencias**

**Departamento de Ciencias Básicas**



**Identificación de factores de mal pronóstico a través de algoritmos de aprendizaje automático en adultos mayores con ataque cerebro vascular isquémico en la región de Ñuble.**

Antonia Belén Benavente Rubio

**Tesis para optar al grado de Magíster en Ciencias biológicas**

**Profesora Guía** Dr Andrés Rodríguez Morales

**CHILLAN, 2021**

## ÍNDICE

|                            |    |
|----------------------------|----|
| Agradecimientos            | 3  |
| Resumen                    | 4  |
| Introducción               | 5  |
| Objetivos e Hipotesis      | 11 |
| Materiales y Métodos       | 12 |
| Resultados                 | 19 |
| Discusión                  | 33 |
| Referencias bibliográficas | 36 |
| Anexos                     | 40 |

### **AGRADECIMIENTOS**

Al Departamento de Ciencias Básicas y a la Escuela de Enfermería por permitirme articular con este programa de postgrado que me abrió tantas oportunidades.

Al Dr. Carlos Escudero por abrirme las puertas de su línea de investigación cuando estaba perdida.

Al Dr. Eduardo López por su gestión en el Hospital Clínico Herminda Martín y su compromiso en generar evidencia científica.

A la Dra. Camila López y a la enfermera Juanita Venegas por su compromiso en la recolección de datos.

Al enfermero Francisco Márquez por su cálido recibimiento en el Centro adosado de Especialidades del Hospital Clínico Herminda Martín.

A cada uno de los docentes e integrantes del programa de magister por inspirarme y guiar el aprendizaje para hacer investigación de calidad.

Al Dr. Andrés Rodríguez por acompañarme en todo momento, enseñarme tantas cosas, contagiarme de su motivación en la etapa más aguda de la pandemia y, por sobre todo, por siempre creer en mis capacidades, a veces más que yo misma.

A mi querida familia y amigos por su apoyo incondicional que me permitió sacar adelante este grado académico.

A mi hijo Pascual por acompañarme siempre, por ser un niño maravilloso que me hizo buscar las mejores oportunidades para los dos.

## RESUMEN

**Introducción:** El Ataque Cerebro Vascular (ACV) es un problema crítico de salud pública que se considera la segunda causa de muerte en Chile y tiene una alta incidencia en la región de Ñuble. Esta patología genera un deterioro severo hasta en un 30% de los afectados, lo que involucra rehabilitación y dependencia grave de terceros. La heterogeneidad que se presenta entre los pacientes con diagnóstico de ACV dificulta la estimación del pronóstico y aún existe a nivel local un escaso desarrollo de modelos predictivos que permitan avanzar en dicha materia. **Objetivo:** Identificar factores de mal pronóstico en adultos mayores con diagnóstico de accidente cerebrovascular isquémico a través de algoritmos de aprendizaje automático. **Metodología:** Se analizaron los parámetros sanguíneos en las primeras 24 horas de manera retrospectiva de 123 adultos mayores tratados en el Hospital Clínico Herminda Martín de Chillán con diagnóstico de Ataque cerebrovascular isquémico, mediante algoritmos de Machine Learning como Análisis de Grupos Jerárquicos (HCA) y Análisis de Componente Principal (PCA). El estudio fue aprobado por el respectivo comité ético científico. **Resultados:** Mayores niveles de eritrocitos (*loading*; 0.43), y menores niveles de colesterol (*loading*;-0.33) fueron identificados como factores de buen pronóstico. Mayores niveles de urea (*loading*;0.41) y creatinina (*loading*;0.29) fueron identificados como factores de mal pronóstico. **Conclusión:** El estudio revela que los algoritmos de machine learning empleados permiten identificar factores de buen y mal pronóstico luego del ACV. Los factores identificados significan información relevante que permitirá mejorar futuros modelos predictivos desarrollados a nivel local, los que serán foco de futuro desarrollo.

## INTRODUCCION

### ***Antecedentes Epidemiológicos***

El Ataque Cerebro Vascular (ACV), se considera la primera causa de muerte tanto en Chile como en la región de Ñuble, donde impacta su alta tasa de mortalidad y letalidad (Lavados et al., 2021). Según cifras entregadas por la Organización Mundial de la Salud (OMS), 15 millones de personas sufren un ACV, de las cuales 5 millones quedan permanentemente discapacitados (Johnson et al, 2016). En Chile, el Ministerio de Salud (MINSAL) ha indicado que existen más de 24.000 casos nuevos por año, lo que significa 70 casos nuevos por día, con una tasa de mortalidad para el año 2015 de 47,1 por cada 100.000 habitantes (MINSAL 2020), dicha tasa es similar a Colombia (54,1 por 100.000 habitantes) o Argentina (52,7 por 100.000 habitantes), pero es drásticamente más alta que las tasas de Estados Unidos (30,4 por 100 000 habitantes) o Canadá (2,7 por 100.000 habitantes) (Vos et al, 2019). Además, esta patología representa la segunda causa de mortalidad prematura en Chile, alcanzando 2.310 defunciones al año 2013 en personas entre 30 y 69 años y se constituye como la primera causa específica de años de vida saludables (AVISA) perdidos por discapacidad y muerte prematura en mayores de 74 años (Minsal 2020).

En Chile, durante las últimas décadas la población de adultos mayores ha ido en aumento, alcanzando un 12% de la población total al año 2019 (SENAMA, 2019). Al año 2035 se proyecta un importante aumento en dicho grupo etario, que alcanzará el 18,9% de la población, según estimaciones del censo 2017. (INE 2019). El ACV se presenta con mayor frecuencia en dicho grupo poblacional, el cual tiene una esperanza de vida estimada de 85 años al año 2045, lo que conllevará a un aumento en la incidencia del ACV y sus respectivas secuelas (Amarenco et al, 2021). Con respecto a esto último, el ACV se considera la primera causa de discapacidad física y cognitiva en el adulto, generando deterioro severo hasta en

un 30% de los afectados, lo que involucra rehabilitación y dependencia grave de terceros (Westerlind, 2020). Dentro de este evento patológico vascular encontramos dos tipos, el ACV hemorrágico que corresponde a una rotura en un vaso sanguíneo cerebral que da origen a una hemorragia y el ACV isquémico (ACVi), trastorno neurovascular caracterizado por la reducción o bloqueo completo del flujo sanguíneo al cerebro (Li et al, 2020). Este último, representa el 72% de los eventos cerebrovasculares anuales en Chile con 11.546 casos estimados al año 2017, motivo por el cual será foco de interés en el presente estudio (Lavados, 2020). Por los antecedentes anteriormente señalados, es que el ACV se constituye como un problema crítico de salud pública.

### ***Manejo del Ataque Cerebrovascular Isquémico en Chile***

Las características epidemiológicas anteriormente descritas han focalizado diversas políticas de salud pública destinadas a dicho evento, es así como el ACVi en personas de 15 años y más, se constituye como la patología N° 37 de las Garantías Explícitas en Salud (GES). Esto garantiza el acceso, la calidad, la oportunidad y la protección financiera para las prestaciones médicas que implica dicha patología, asegurando el diagnóstico, tratamiento, seguimiento y rehabilitación (Minsal, 2020). Las medidas están enfocadas principalmente en el diagnóstico y tratamiento, pero aún existen deficiencias en torno al seguimiento y a la rehabilitación, lo que implica una deficiencia en la prevención terciaria, que se define como el conjunto de acciones destinadas a disminuir el progreso de la enfermedad (Stinear et al, 2020).

Además, es preciso señalar que esta patología presenta gran diversidad etiológica y una limitada disponibilidad de métodos diagnósticos de alta precisión, lo que dificulta el diagnóstico, tratamiento y pronóstico individualizado (Zuo et al.,2020).

Dicho esto, se presenta la inquietud de mejorar dicha realidad a través de una investigación que permita identificar el valor pronóstico de las variables clínicas del usuario, con enfoque multifactorial e individual.

### **Secuelas motoras y cognitivas post ataque cerebrovascular isquémico**

Las secuelas neurológicas más importantes luego de un ACV se producen en los primeros 6 meses (Lavados, 2007). Las discapacidades físicas más importantes observadas son las motoras, disfagia y afasia (Sennfalt, 2021). Con relación a las secuelas cognitivas, se observa un deterioro que abarca un amplio espectro de cambios cognitivos y conductuales, caracterizados por un deterioro de la atención, comprensión, expresión, comunicación, interacción social y memoria (Luna-Matos et al.,2007).

### ***Pronóstico del Ataque Cerebrovascular Isquémico***

La heterogeneidad que se presenta entre los pacientes con diagnóstico de ACVi dificulta la estimación del pronóstico (Han et al, 2019). De acuerdo con el protocolo disponible en la Guía Clínica para el manejo del ACV del ministerio de salud, se utiliza la escala denominada *National Institutes of Health Stroke Scale* (NIHSS), la que permite evaluar cuantitativamente la severidad del ACV y deterioro neurológico en el usuario de acuerdo con su estado de conciencia, ubicación temporal, lenguaje, fuerza en las extremidades, visión sensibilidad y otras (Yaghi et al, 2017). De acuerdo con el puntaje obtenido se diferencian 4 niveles; leve (0-5 pts), moderado (6-10 pts), severo (16-19 pts) y muy severo (20 y más pts), lo que permite apoyar la toma de decisiones con respecto al pronóstico y tratamiento (Zhao, 2018). Sin embargo, recientes estudios han establecido que la escala NIHSS no es suficiente por sí sola para determinar el deterioro neurológico y pronóstico de cada individuo (Abzhandadze et al, 2020). Adicionalmente, a pesar de que en la guía clínica se enfatiza la necesidad de realizar la evaluación neurológica con la escala NIHSS (Anexo 1) con la mayor

prontitud posible, el estudio ÑANDU realizado en los años 2015-2017 en Ñuble señala que sólo un 7,4% de los usuarios recibieron dichas evaluaciones en su primer contacto con el servicio de urgencias y un 15,6% fue evaluado pasado las 24 horas desde el inicio de síntomas (Lavados et al, 2021).

Otra escala ampliamente utilizada para evaluar el ACV, es la denominada *modified Rankin scale* (mRs), la que permite medir el grado de dependencia para realizar las actividades de la vida diaria en los usuarios luego de un episodio de ACV (Broderick, 2017). En dicha evaluación se establece un puntaje del 0 al 6, que va desde la ausencia de síntomas, abordando niveles de dependencia moderados y severos, hasta la muerte (Anexo 2). La escala mRS se ha utilizado en diversos estudios predictivos para establecer el pronóstico funcional de los usuarios (Chen, 2021), (Cramer, 2021), (Lattanzi, 2020).

Con objeto de lograr una mayor precisión en el diagnóstico, pronóstico y/o tratamiento del ACV, se han estudiado diversos biomarcadores que predicen un mal pronóstico en el ACV isquémico, pero todavía no se ha podido identificar uno que tenga suficiente valor predictivo como para ser incorporado en la práctica clínica (Alfieri et al, 2020). Diversos estudios se han centrado en generar modelos predictivos que permitan estimar el pronóstico y resultado clínico post ACV, pero lograr un modelo que sea aceptado ampliamente para ser incorporado a los protocolos clínicos continúa como un desafío (Lin et al, 2020). Por otro lado, es necesario entrenar modelos a partir de los parámetros y características de la población local, para así evitar sesgos en los algoritmos utilizados en salud (Wiens, 2020). La predicción del resultado clínico post ACV con tiempo suficiente para intervenir y poder mejorar la rehabilitación de los usuarios sigue manteniéndose como un desafío para la salud pública de Chile (Lanas et al, 2020). Además, aún existe a nivel local un escaso desarrollo de modelos predictivos que permitan avanzar en medicina de precisión.

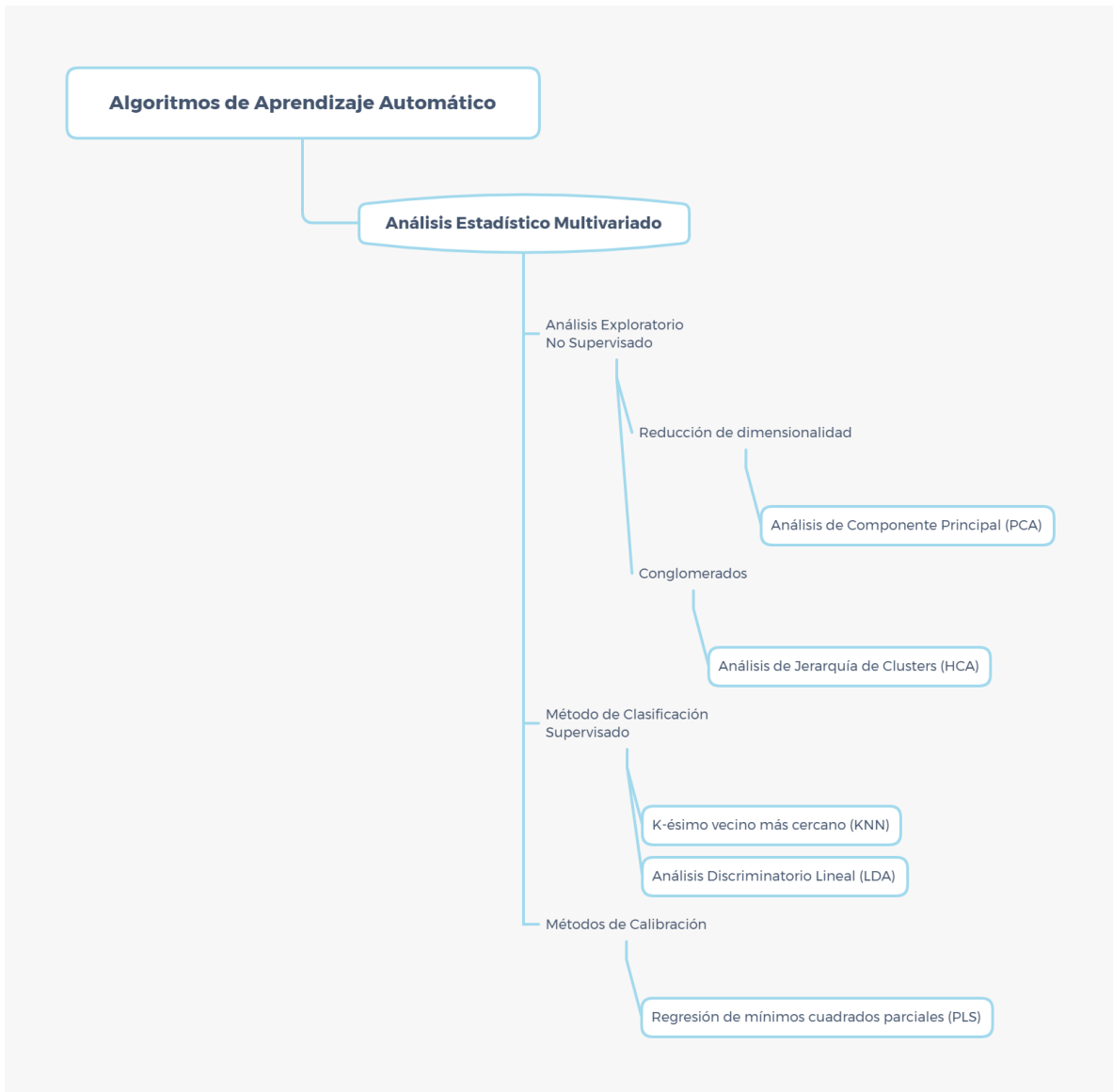
Actualmente, según el protocolo clínico de ACV utilizado en el sistema de salud de Chile, se realizan una serie exámenes sanguíneos a los usuarios afectados por ACV al llegar al



servicio de urgencias y durante su hospitalización. Dichos exámenes incluyen perfil sanguíneo, perfil lipídico, perfil hepático, perfil renal, electrolitos plasmáticos, perfil glicémico, perfil inflamatorio, perfil de coagulación, lo que significa un importante consumo de recursos sanitarios. Debido a que estos valores son realizados de manera rutinaria, son una fuente importante de variables que pueden ser extraídas de los bancos de estadística de los centros de salud, para luego identificar su valor pronóstico a través de algoritmos de aprendizaje automático, campo de la inteligencia artificial.

### ***Algoritmos de aprendizaje automático aplicado en salud***

Dado que los fenómenos fisiopatológicos son de origen multifactorial, el aprendizaje automático se vuelve una herramienta de gran utilidad que permite analizar un gran número de variables e identificar su valor pronóstico para un determinado evento (Kogan et al, 2020). Este tipo de análisis, permite el procesamiento de una gran cantidad de datos, reduciendo su dimensionalidad para identificar aquellas variables de mayor importancia pero sin perder información relevante (Peng et al, 2020). El aprendizaje automático incluye diferentes algoritmos supervisados como no supervisados, tales como HCA, PCA, KNN y PLS (Caron et al, 2020), las que serán definidas en el apartado de metodología. El Análisis de Componente Principal (PCA) se ha identificado como un análisis que permite identificar el valor pronóstico de ACV en los usuarios de Chile y que revela correlaciones que no son observables mediante la estadística clásica (Escudero et al, 2021). Es necesario entonces, profundizar en el estudio de algoritmos de aprendizaje automático para pronosticar el resultado clínico de los usuarios con diagnóstico de ataque cerebrovascular isquémico. Disponer de modelos predictivos que respondan a las características individuales de la población local, resulta imprescindible para lograr un pronóstico preciso e individualizado, pero aún permanece como un desafío para la salud de Chile.



**Algoritmos de aprendizaje automático.**

**Hipótesis:**

El uso de algoritmos de aprendizaje automático permite identificar factores de mal pronóstico en adultos mayores con diagnóstico de accidente cerebro vascular isquémico en la región de Ñuble.

### **Objetivos:**

#### **Objetivo general:**

Identificar factores de mal pronóstico en adultos mayores con diagnóstico de accidente cerebrovascular isquémico a través de algoritmos de aprendizaje automático.

#### **Objetivos específicos:**

1. Describir las características sociodemográficas y clínicas de la muestra en estudio según pronóstico.
2. Identificar la similitud entre las variables clínicas en los diferentes grupos, según pronóstico
3. Identificar las variables clínicas que tienen mayor importancia en el pronóstico de los usuarios con diagnóstico ACV isquémico.

## MATERIALES Y MÉTODOS

### METODOLOGÍA:

#### Tipo de diseño:

El estudio posee un enfoque cuantitativo, analítico, transversal, observacional y retrospectivo.

#### Población de estudio:

Se consideró el total de adultos mayores, hombres y mujeres que fueron tratados en el servicio de neurología a causa de un ataque cerebro vascular isquémico en el Hospital Clínico Herminda Martín de Chillán. Esta cifra es de aproximadamente 1200 usuarios al año. Considerando solo los datos clínicos de los años 2018 y 2019. No se utilizaron los datos correspondientes al año 2020, ya que por la pandemia por covid-19 se aceleró el proceso de alta médica, donde se aplica la escala mRs.

#### Muestra:

Considerando una población de 1100 usuarios con diagnóstico de ACV isquémico por año, con un nivel del confianza del 95% y precisión del 5%, el tamaño muestral calculado fue de:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{d^2 * (N - 1) + Z_{\alpha}^2 * p * q} = \frac{((1100) * (1.96)^2 * (0.1) * (0.9))}{(0.05^2 * (1100 - 1) + (1.96)^2 * (0.1) * (0.9))} = 123$$

**N** = Total de la población = 1100

**Z $\alpha$**  = 1.96 (para nivel de confianza del 95%)

**p** = proporción esperada = 10% = 0.1

**q** = (1-p) = 1-0. = 0.9

**d** = precisión = 5%

Dicho tamaño muestral (123 usuarios) fue recolectado con éxito.

### **Matriz de análisis**

Considerando un análisis de 23 variables clínicas estudiadas en una muestra de 123 usuarios, se analizó una matriz correspondiente a un total de 2.829 datos, cantidad que justifica el uso de algoritmos de aprendizaje automático que permite hacer reducción de dimensionalidad, como el Análisis de Componente Principal (PCA).

### **Unidad de análisis:**

Adulto mayor ingresado en el servicio de neurología del HCHM con diagnóstico de ACV isquémico.

### **Criterios de elegibilidad:**

#### **Criterios de Inclusión:**

- Usuario mayor o igual a 65 años.
- Usuario con diagnóstico de accidente cerebrovascular isquémico.
- Usuario hospitalizado en el servicio de Neurología del Hospital Clínico Herminda Martín (HCHM) durante los años 2018 y 2019

#### **Criterios de Exclusión:**

- Usuario sin batería de exámenes disponible en ficha clínica.
- Usuario trasladado a otro hospital.
- Usuario con antecedentes previos de enfermedades cerebro vasculares (ECV).
- Usuario con secuelas neurológicas previas.

**Técnica de muestreo empleada:** Los usuarios fueron seleccionados con una técnica de muestreo aleatorio simple entre los usuarios de la población que cumplían con los criterios de inclusión previamente descritos.

**Descripción de instrumentos recolectores:**

Se confeccionó manualmente una base de datos anónima con las variables de interés (Anexo 1) a partir de la información que se extrajo desde la epicrisis médica disponible y el sistema en línea de exámenes Syslab. Los parámetros sanguíneos analizados, corresponden a la primera toma de exámenes sanguíneos en el servicio de urgencia del HCHM, todos ellos medidos en menos de 24 horas a partir del evento ACV. Se incluyeron parámetros sanguíneos correspondientes al hemograma, perfil renal, perfil hepático, electrolitos plasmáticos y otros parámetros del perfil bioquímico.

**Procedimiento para recolección de datos:**

Se realizó la solicitud al director del Hospital Clínico Herminda Martin para acceder a la fuente de información, luego se presentó el estudio al Comité Ético Científico de dicho establecimiento. Una vez aprobado, se accedió al sistema Syslab y epicrisis médica de cada usuario de donde se extrajo los datos en función de los objetivos de estudio. No se trabajó con datos sensibles de los usuarios, por lo que cada uno fue identificado con un número, sin trabajar con nombres, apellidos ni RUT.

**Aspectos éticos:**

Se realizó mediante los protocolos establecidos por la Comisión Ministerial de Ética de la Investigación en Salud (CMEIS) y teniendo en cuenta la Ley 20.120, Ley 20.584, la presentación del proyecto para evaluación por el Comité Ético Científico (CEC) del Hospital

Clínico Herminda Martín (HCHM), acreditado 24 de diciembre, 2020 según Resolución Exenta N° 25139/2020 de la SEREMI de Salud de la Región de Ñuble. El presente estudio fue aprobado por el CEC del HCHM mediante el ordinario N°31/N°20/20. Si bien todas las investigaciones que involucran personas o sus datos requieren consentimiento informado, considerando que los datos eran retrospectivos, correspondientes a los años 2018 y 2019, que la cantidad de usuarios a abordar era elevada, y al alto porcentaje de pacientes fallecidos o con daño cognitivo dependiente, el poder acceder a solicitar un consentimiento informado individualizado era de gran dificultad. Por ello se solicitó una dispensa del consentimiento informado al CEC argumentando una relación riesgo-beneficio favorable para el proyecto, y que según las consideraciones emitidas por la CMEIS hacen referencia a: i) la necesidad de utilizar datos personales, porque no hay alternativas válidas para el estudio ni se podría utilizar información anonimizada; ii) la importancia de la investigación; iii) el bajo riesgo del uso indebido de los datos personales y las garantías para protegerlos. Una vez aprobado, se accedió a la ficha clínica de los sujetos incluidos en esta investigación. (Anexo 2)

.

#### **Custodia de los datos:**

El/La profesional encargado/a ingresó la información a una base de datos secundaria, la cual es de acceso restringido. El custodio de la información será la ejecutante, Antonia Bevnavente Rubio y el profesor tutor, Andrés Rodríguez Moralez, del proyecto durante todo el periodo de la investigación, con un máximo de 5 años, en una base virtual con acceso cifrado. Dicha base de datos no será compartida con terceros y será eliminada luego de 5 años. De surgir un análisis post-hoc, esta deberá nuevamente ser autorizado

#### **Procesamiento y presentación gráfica de los datos:**

Todos los análisis se realizaron con el software estadístico Pirouette® (Infometrix., Inc.), mientras que los gráficos fueron desarrollados en el programa PRISM® (GraphPad, Inc).

**Pre-procesamiento:**

Una vez obtenida la base de datos con las variables de interés, se verificó que todos los usuarios tuvieran un valor numérico para cada una de las variables incluidas. La estandarización de los datos se realizó utilizando el pre-procesamiento autoescala para el PCA y pre-procesamiento centrado en la media para el caso de HCA. Las variables categóricas como buen y mal pronóstico se transformaron a códigos numéricos dicotómicos (1/0). Los datos se ingresaron a la matriz de procesamiento, en la cual se eliminaron los datos que presentaron valores atípicos según distancia de Mahalanobis.

$$Xs = \frac{X - media}{ds}$$

*Formula Autoescala. X corresponde a la variable, ds a la desviación estándar.*

**Varianza vs variabilidad**

La desviación estándar, s, la raíz cuadrada de la varianza, es quizás una medida más familiar de la dispersión en una serie de valores. Al utilizar el preprocesamiento autoescalado o el centrado en la media, la distinción entre variación y varianza es discutible; ya que dichas opciones de preprocesamiento resta la media de cada variable, de modo que las dos cantidades sólo difieren en un factor de n-1. Sin embargo, si no preprocesa sus datos o escala el rango o la varianza, debe tener en cuenta la distinción variación/varianza. En el presente estudio se utilizará el término varianza ya que todos los análisis se realizaron con preprocesamiento autoescala o centrada en la media. Dicha varianza representa la desviación o separación de los datos, por lo que los componentes principales o variables que tienen una mayor varianza son aquellos que tendrán una mayor efecto en la separación y agrupación espontánea y, por lo tanto, aquellas de mayor importancia para el estudio.



## Procesamiento de Datos:

### Modified Rankin Scale (mRS)

Los usuarios fueron clasificados en 2 grupos según su puntaje en la escala Modified Rankin Scale (mRS) aplicada en el alta médica (Anexo 3), los usuarios con puntaje  $mRS \leq 2$  fueron categorizados en el grupo “buen pronóstico” y aquellos usuarios con puntaje  $mRS \geq 3$  fueron categorizados en el grupo “mal pronóstico” (Wilson et al. 2002). La variable “pronóstico” no fue parte del procesamiento de los datos, se mantuvo únicamente como variable categórica.

### Análisis de Jerarquía de Clusters (HCA)

Para identificar la similitud entre las variables clínicas en los diferentes grupos pronósticos, se utilizó el Análisis de Jerarquía de Clusters (HCA), que permitió agrupar las variables logrando la mayor homogeneidad entre los grupos formados espontáneamente y la mayor diferencia entre ellos. Fue útil para realizar un primer filtro que identificó qué variables fueron las que se comportaron de manera similar en una cierta categoría como, por ejemplo, aquellos usuarios de mal pronóstico. La relación entre variables se determinó mediante la distancia de tipo Euclidiana, mientras que la generación de conglomerados según similitud se realizó mediante el cálculo de unión incremental. Las agrupaciones formadas por el algoritmo HCA se ilustran mediante un dendrograma, representación gráfica que permite apreciar con claridad las relaciones de agrupación entre los datos.

$$d_{ab} = \sqrt{\sum_i^m (x_{aj} - x_{bj})^2}$$

*Formula Distancia Euclidiana*

$$d_{AB} = C \left( \frac{(n_A + n_C)d_{AC}^2 + (n_B + n_C)d_{BC}^2 - n_C d_{AB}^2}{n_A + n_B + n_C} \right)^{1/2}$$

*Formula de Unión Incremental*

### **Análisis de componente principal (PCA)**

Con el fin de identificar las variables clínicas de mayor importancia en el pronóstico de los usuarios con diagnóstico ACV isquémico, se utilizó el algoritmo de Análisis de Componente Principal (PCA), el cual permitió agrupar usuarios de acuerdo a la similitud de sus patrones clínicos e identificar los componentes principales más importantes que diferencian el grupo de usuarios de buen y mal pronóstico. Dichos componentes principales pueden ser un pequeño número de factores que explican la varianza observada en un número mucho mayor de variables clínicas medidas en los pacientes con diagnóstico de ACV, identificando correlaciones ocultas en datos analizados por estadística no multivariada. En efecto, permite reducir la dimensión de los datos sin perder información relevante. Las variables se consideraron significativas según la regla de Malinowsky, con un factor de carga >0,25.

## **RESULTADOS**

### **1. Características sociodemográficas y clínicas de la muestra en estudio**

Las características de la muestra están resumidas por grupo pronóstico en la tabla 1. Se incluyeron n=123 usuarios; el 55.74% (n=68) correspondió al grupo de buen pronóstico, mientras que el grupo de mal pronóstico alcanzó el 44.26% (n=54). De acuerdo a los diferentes parámetros sanguíneos analizados, se encontró diferencia estadísticamente significativa según pronóstico en el valor promedio de plaquetas (p<0.001), colesterol total (P<0.001), urea (p=0.0073), nitrógeno uréico (p=0.012) y glicemia (p<0.001).

**Tabla 1.** Características de los usuarios según pronóstico

---

|                   | Buen pronóstico <sup>a</sup><br>(Media ± DS) | Mal pronóstico<br>(Media ± DS) | P value          | Valor normal examen <sup>b</sup> |
|-------------------|--|--------------------------------|------------------|----------------------------------|
| n (%)             | 68 (55.74)                                   | 54(44.26)                      |                  | -                                |
| Edad              | 75.27 ± 9.36                                 | 77.94 ± 9.23                   |                  | -                                |
| Eritrocitos       | 4.54 ± 0.49                                  | 4.63 ± 1.68                    |                  | 3.80-5.20 x10 <sup>6</sup> μL    |
| Hematocrito       | 40.36 ± 3.81                                 | 39.39 ± 5.06                   |                  | 35.0 - 47.0 %                    |
| Hemoglobina       | 13.56 ± 1.93                                 | 13.25 ± 1.78                   |                  | 11.6 - 15.5 g/dL                 |
| Plaquetas         | 237.29 ± 56.20                               | 217.17 ± 72.37                 | <b>&lt;0.001</b> | 140-440x10 <sup>3</sup> μL       |
| Leucocitos        | 9.25 ± 3.98                                  | 8.95 ± 2.57                    |                  | 4-10 K/ μL                       |
| Colesterol total  | 164.61 ± 42.03                               | 136.59 ± 36.64                 | <b>&lt;0.001</b> | <200 mg/dL                       |
| HDL               | 40.13 ± 10.32                                | 41.53 ± 11.72                  |                  | >35 mg/dL                        |
| LDL               | 99.44 ± 38.40                                | 73.85 ± 35.88                  | <b>&lt;0.001</b> | 50-160 mg/dL                     |
| Trigliceridos     | 125.75 ± 40.06                               | 113.05 ± 81.09                 |                  | <150 mg/dL                       |
| Creatinina        | 0.94± 0.34                                   | 1.16 ± 1.01                    |                  | 0.60-1.20 mg/dL                  |
| Urea              | 38.99 ± 14.93                                | 48.46 ± 21.61                  | <b>0.0073</b>    | 6-40 mg/dL                       |
| Nitrógeno Ureico  | 18.15 ± 7.02                                 | 22.64 ± 10.09                  | <b>0.012</b>     | 7-25 mg/dL                       |
| Cloro             | 103.41 ± 3.44                                | 100.98 ± 4.21                  |                  | 98-107 mEq/L                     |
| Sodio             | 136.08 ± 2.82                                | 134.87 ± 3.31                  |                  | 136-145 mEq/L                    |
| Potasio           | 4.02 ± 0.52                                  | 4.04 ± 0.48                    |                  | 3.5-5.10 mEq/L                   |
| HbA1c             | 6.45 ± 1.48                                  | 6.97 ± 2.30                    |                  | 4.4-6.0 %                        |
| Glicemia          | 133.51 ± 41.23                               | 171.59 ± 103.27                | <b>&lt;0.001</b> | 70-105 mg/dL                     |
| PCR               | 12.52 ± 22.18                                | 20.31 ± 30.47                  |                  | 0-8.2 mg/dL                      |
| INR               | 1.19 ± 0.37                                  | 1.45 ± 1.41                    |                  | 0.9-1.3                          |
| TTPK              | 28.34 ± 6.99                                 | 29.15 ± 10.18                  |                  | 22.6-34.1 s                      |
| Tiempo protombina | 13.25 ± 3.95                                 | 17.84 ± 18.25                  |                  | 11-13.5 s                        |
| Troponina I       | 0.47 ± 3.08                                  | 0.21 ± 0.56                    |                  | <0.040 μg/L                      |

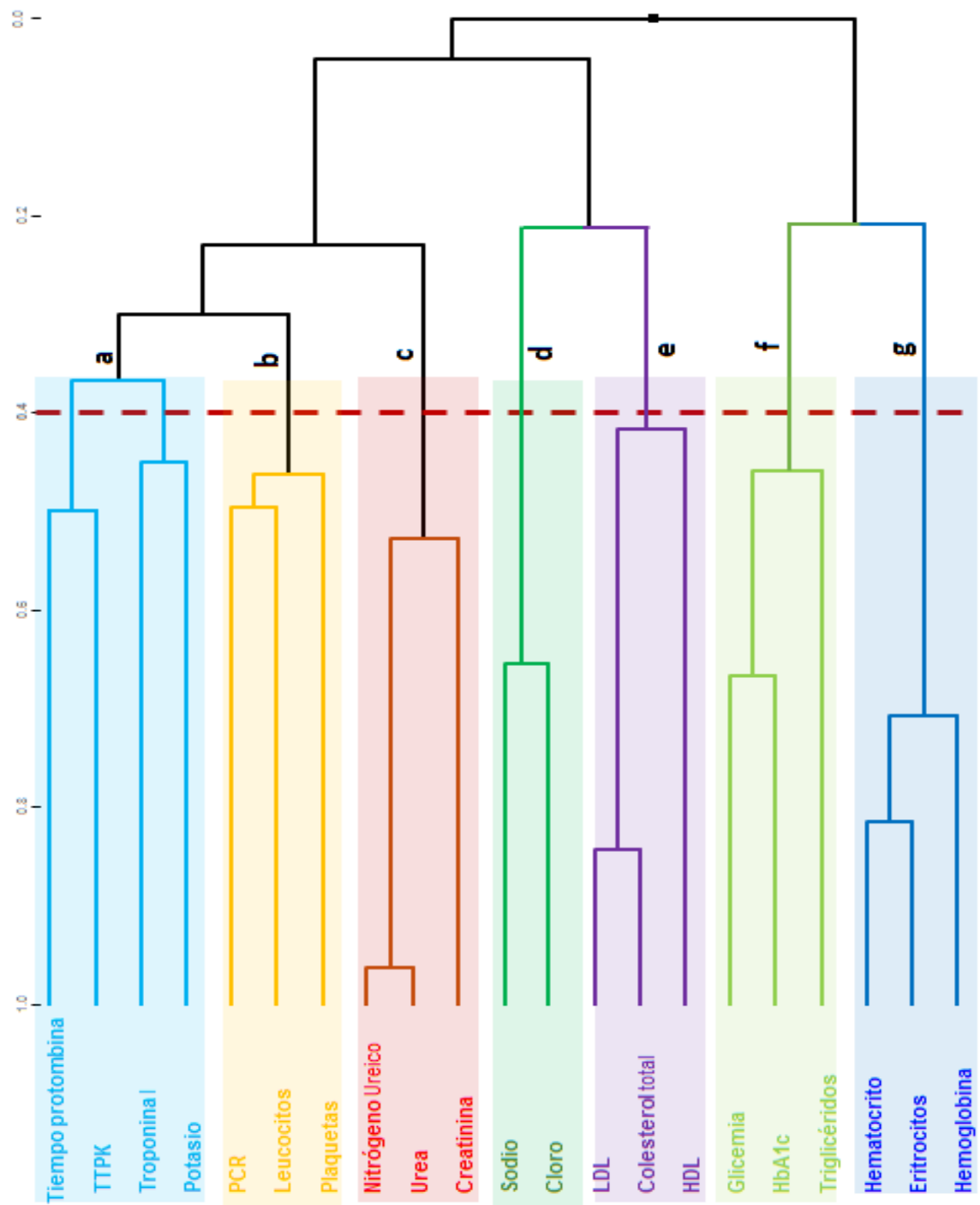
**a** Buen pronóstico: mRS ≤ 2; Mal pronóstico: mRS ≥ 3 (Wilson et al. 2002)

**b** Laboratorio Clínico Hospital Clínico Herminda Martin

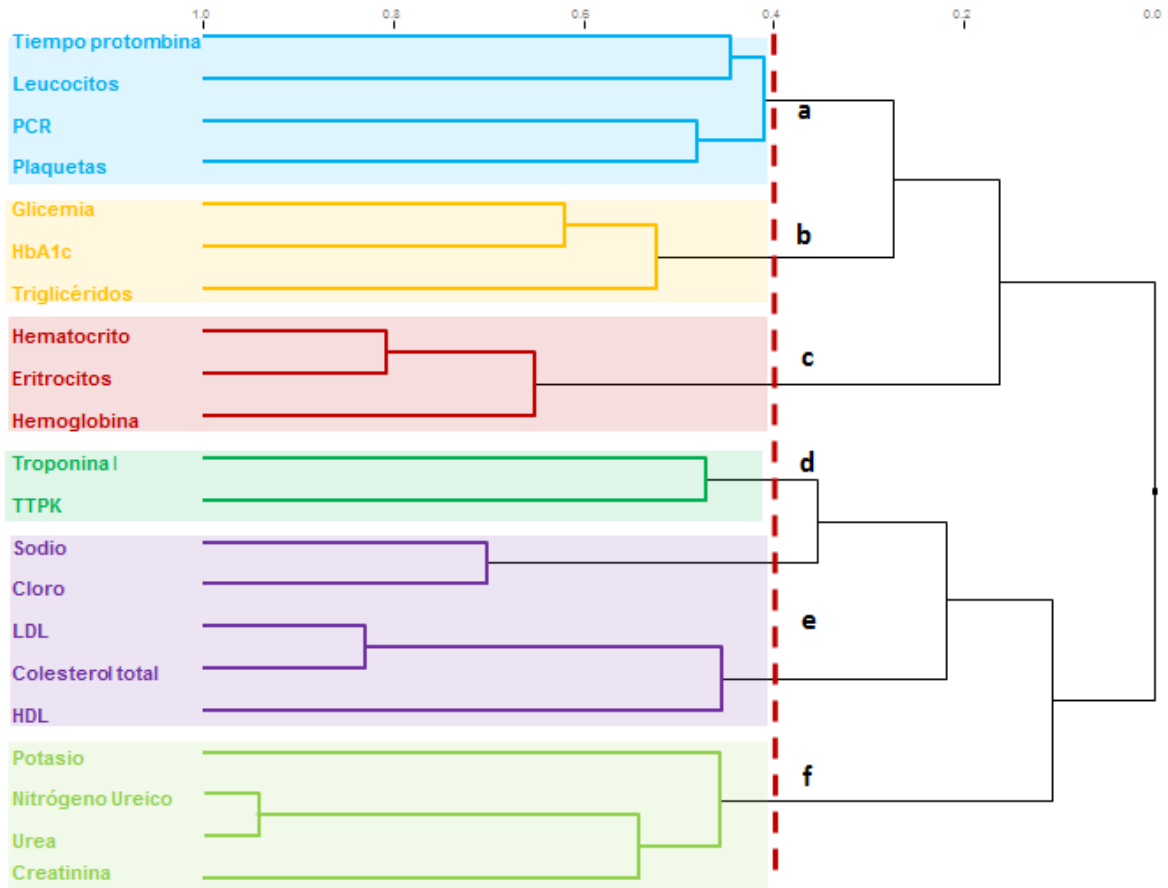
c P value según prueba t de Student

## 2. Similitud de variables clínicas entre usuarios según HCA

Los conglomerados o *clusters* formados espontáneamente en la muestra general, sin separación de grupo pronóstico, están descritas a través del dendograma (figura 1). Se identificaron 7 principales *clusters* analizados con un porcentaje de similitud de 40%: el *cluster* “a” asoció parámetros de coagulación como Tiempo de protrombina, TTPK y potasio. El *cluster* “b” asoció parámetros inflamatorios, incluyendo las variables Proteína C - Reactiva (PCR), leucocitos y plaquetas. El *cluster* “c” asoció parámetros clínicos de función renal; las variables Nitrogeno Ureico, Urea y Creatinina. El *cluster* “d” asoció los niveles de electrolitos plasmáticos correspondientes a Sodio y Cloro. El *cluster* “e” agrupó Colesterol Total, Lipoproteínas de Alta Densidad (HDL) y Lipoproteínas de Baja Densidad (LDL), pertenecientes al perfil lipídico. El *cluster* “f” agrupó los parámetros Glicemia, Hemoglobina Glicosilada (Hb1Ac) y Triglicéridos, asociados al metabolismo y regulación de la glucosa. Finalmente, el *cluster* “g” asoció variables pertenecientes al hemograma; Eritrocitos, Hematocrito y Hemoglobina. Según lo descrito, se observa que las variables se agrupan de un modo concordante con los parámetros asociados a relaciones fisiológicas o fisiopatológicas según corresponda.



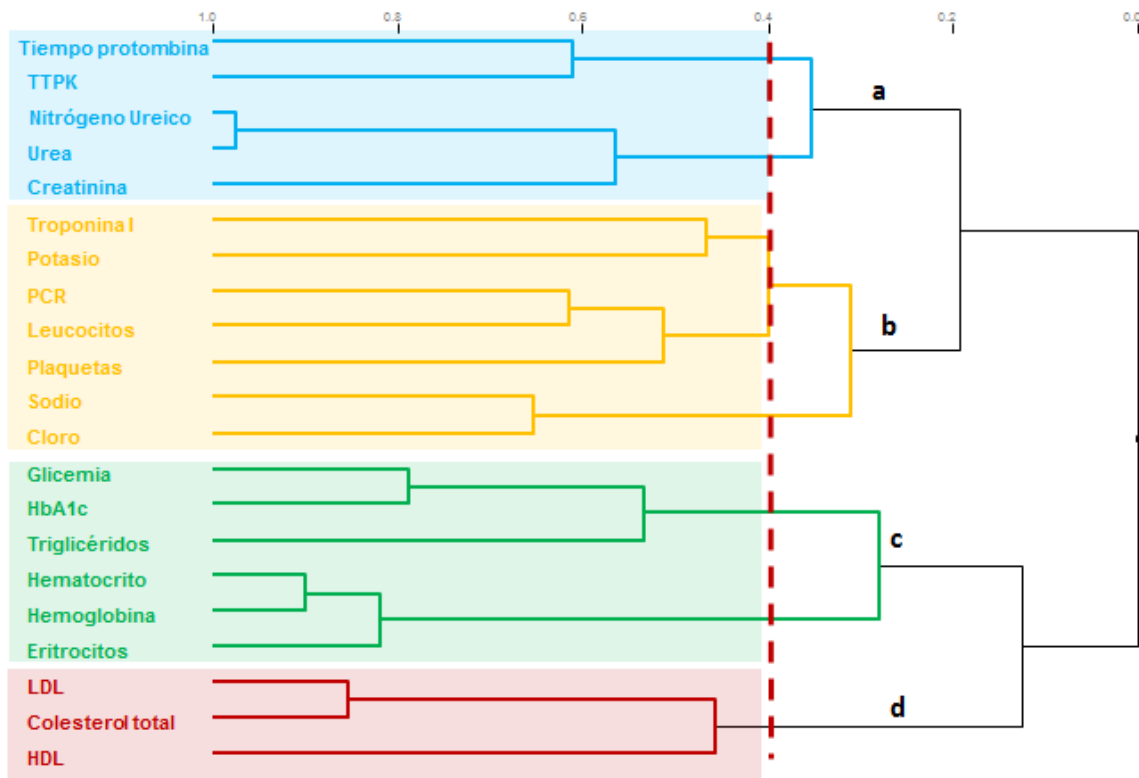
**Figura 1. Similitud de parámetros sanguíneos sin distinción de grupo pronóstico.**  
 Dendograma basado en análisis incremental de jerarquía de clusters (HCA) de similitud de 20 parámetros sanguíneos sin distinción de grupo pronóstico analizado con 40% de similitud.



**Figura 2. Similitud de parámetros sanguíneos entre usuarios de buen pronóstico.** Dendograma basado en análisis incremental de jerarquía de clusters (HCA) de similitud de 20 parámetros sanguíneos correspondiente al grupo de buen pronóstico, analizados con 40% de similitud.

Al activar la marca clase o variable categórica, buen y mal pronóstico, se observan diferencias en la formación de clusters espontáneos. En el Análisis de Jerarquía de Clusters (HCA) del grupo de usuarios categorizados como de buen pronóstico ( $mRs \leq 2$ ) (figura 2), se observa la formación de 6 clusters con 40% de similitud. En este grupo de usuarios de buen pronóstico, el primer cluster “a” asocia parámetros de coagulación (Tiempo de protrombina y Plaquetas) junto a parámetros de carácter inflamatorio, como proteína C-reactiva (PCR) y Leucocitos.

En el grupo de mal pronóstico (figura 3) en cambio, se observa que el cluster “a” agrupa parámetros de coagulación (Tiempo de protrombina y TTPK) con parámetros de función renal (Nitrógeno Ureico, Urea y Creatinina).



**Figura 3. Similitud de parámetros sanguíneos entre usuarios de mal pronóstico.**

Dendrograma basado en análisis incremental de jerarquía de conglomerados o *clusters* (HCA) de similitud de 20 parámetros sanguíneos correspondiente al grupo de mal pronóstico analizado con 40% de similitud.

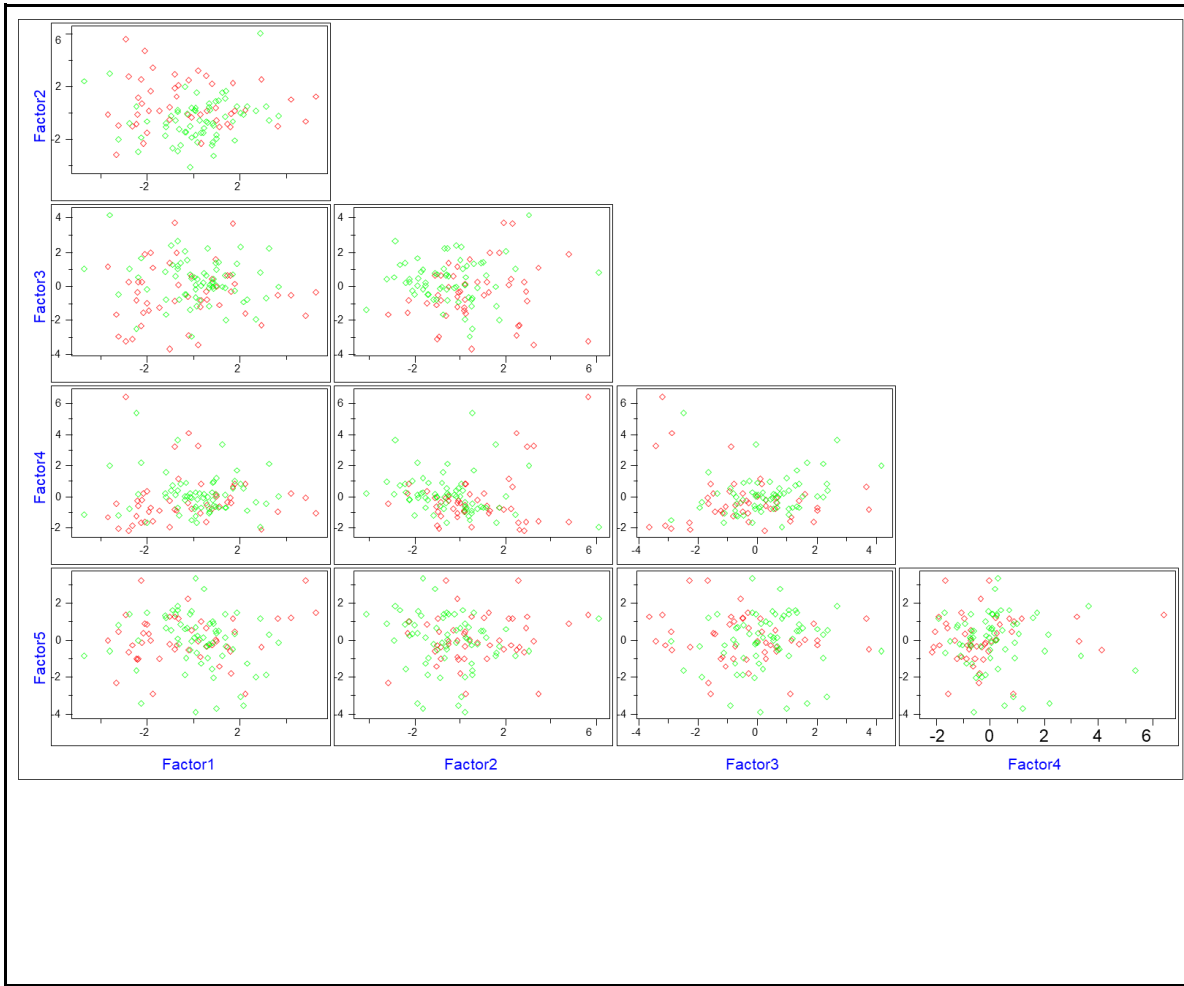
### 3. Variables clínicas de mayor importancia en el pronóstico de los usuarios con diagnóstico ACV isquémico.

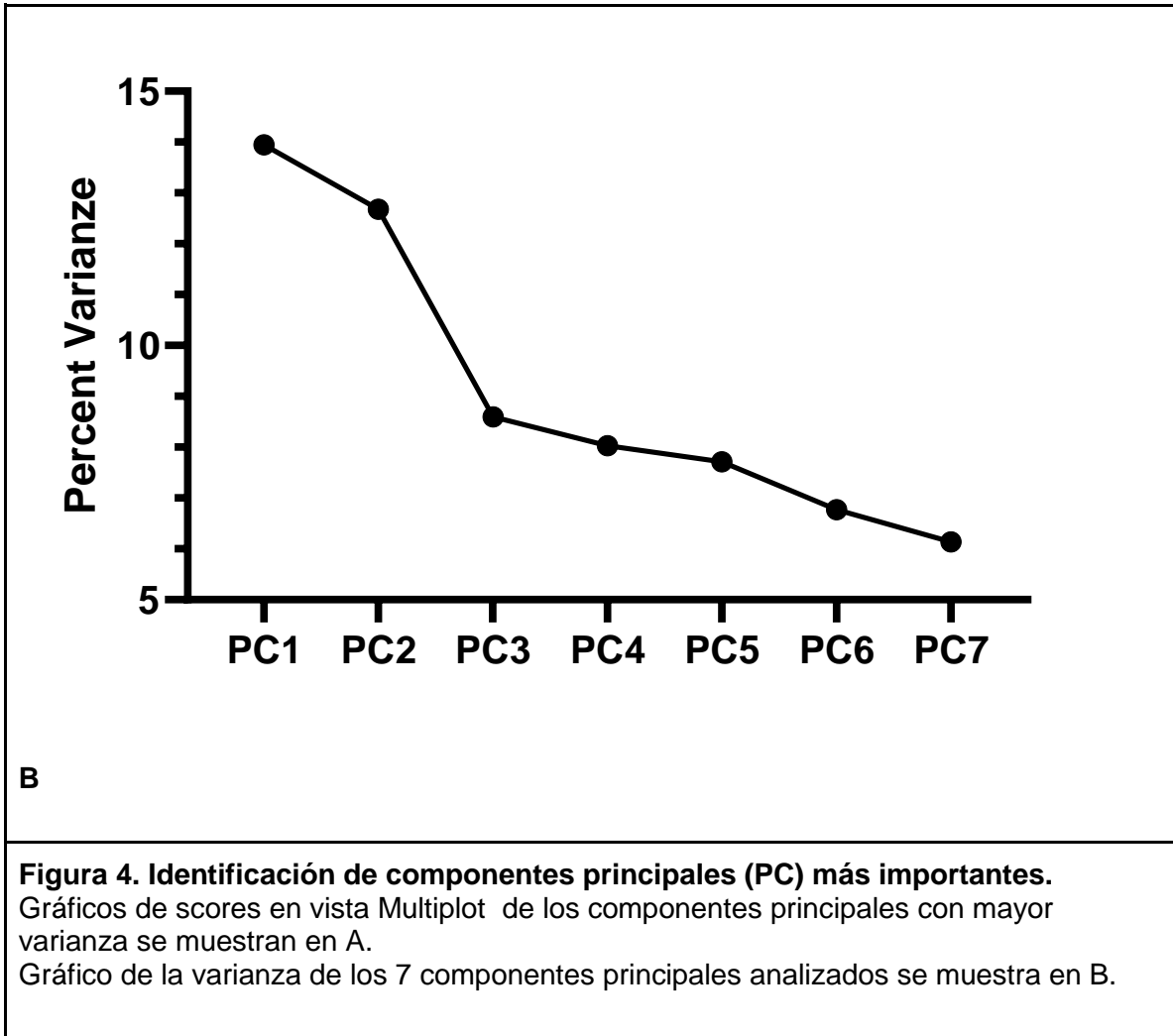
Tal como se describe en el apartado de metodología, PCA es una herramienta útil para

identificar patrones en bases de datos de gran tamaño, sin necesidad de una clasificación previa y conservando la información más relevante. En el presente estudio se utilizó PCA para valorar la importancia de las variables clínicas en el pronóstico del ACV.

Los patrones de dispersión de los usuarios con ACV isquémico según pronóstico se detallan en la figura 4;A. Los usuarios de buen pronóstico ( $mRS \leq 2$ ) se han identificado con color verde y aquellos con mal pronóstico ( $mRS \geq 3$ ) se han identificado con color rojo. Se observa que en la mayoría de los gráficos no existe formación espontánea de clusters que se correspondan con la variable de clase asignada, excepto en el gráfico PC1 vs PC2. Desde el punto de vista de la varianza de los 7 componentes principales analizados (figura 4;B), se observa que los componentes principales 1 y 2, presentan las varianzas más elevadas 13,9% y 12,3% respectivamente. Para los próximos análisis entonces, se seleccionarán los PC1 y PC2, ya que según el método del codo o elbow rule (Kodinariya et al,2013), el punto de quiebre se produce a partir del PC3 (8,4%), seguido de una disminución de la pendiente de la curva de varianza en los siguientes componentes principales





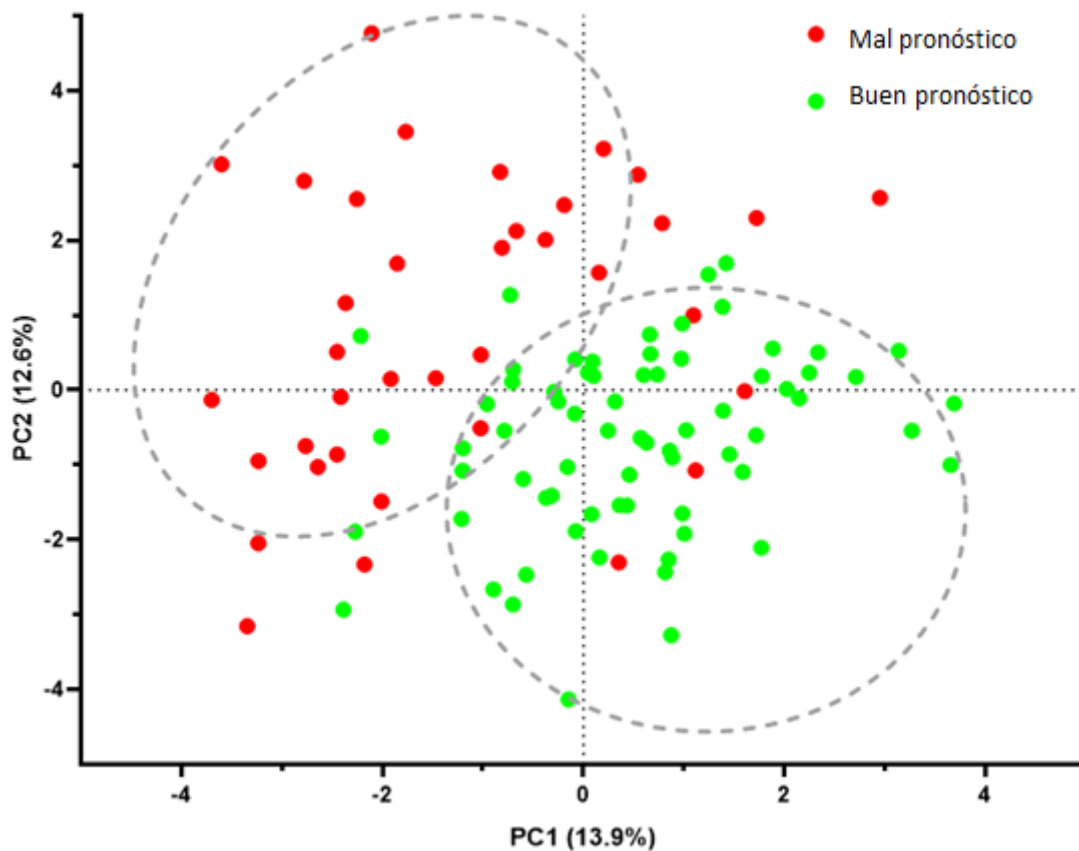


Los primeros PC están graficados en figura 5, los colores representan la variable de clase previamente descrita donde el color verde corresponde al grupo de buen pronóstico y el color rojo al grupo de mal pronóstico.

Al analizar el gráfico de scores (Figura 5), se puede observar una distribución agrupada de manera espontánea, no supervisada, en la que los usuarios de buen pronóstico al alta médica luego de sufrir un ACV isquémico se sitúan predominantemente en el cuadrante inferior derecho, mientras que los usuarios de mal pronóstico se sitúan predominantemente en el cuadrante superior izquierdo. Al analizar el eje del primer componente principal (PC1), el cual describe el 13.9% de la varianza de la matriz de datos, se observa que este distribuye

a los usuarios de buen pronóstico predominantemente hacia la derecha del punto cero, es decir donde se sitúan los valores positivos. Por otro lado, los usuarios de mal pronóstico se distribuyen a la izquierda del punto cero (valores negativos del PC1).

El segundo componente principal (PC2) el cual describe el 12.6% de la varianza de la matriz de datos, distribuye a los usuarios de buen pronóstico predominantemente hacia arriba del punto cero, es decir donde se ubican los valores positivos. En adición, los usuarios de mal pronóstico se distribuyen hacia abajo del punto cero, donde se sitúan los valores negativos del PC2.



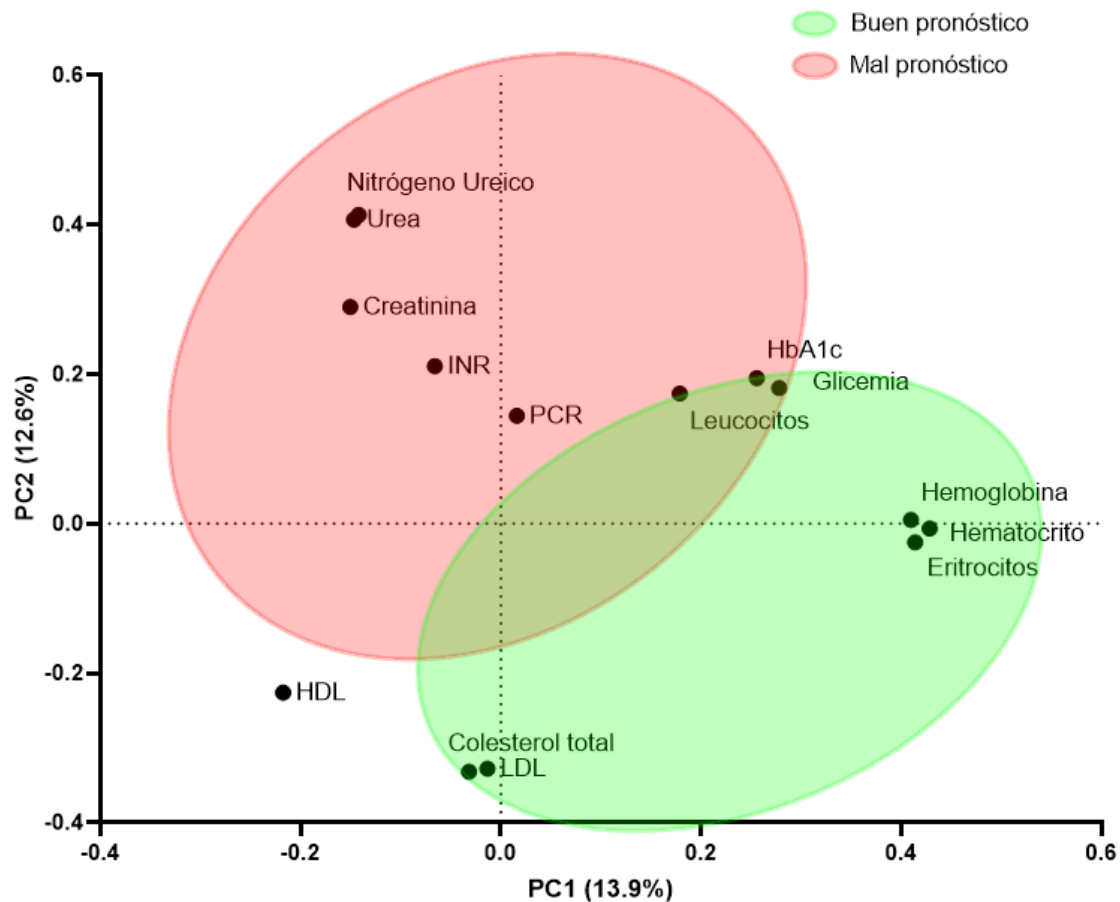
### **Figura 5. Distribución espontánea de usuarios según PCA**

El gráfico de puntuaciones o *scores* muestra la distribución espontánea y formación de conglomerados o *clusters* de usuarios, proyectados en las nuevas dimensiones descritas por el componente principal 1 y 2. Los puntos verdes corresponden a los usuarios de buen pronóstico y los puntos rojos corresponden al grupo de mal pronóstico. En paréntesis está expresado el porcentaje de varianza de cada componente principal.

Para identificar las variables más importantes en la separación de los grupos de buen y mal pronóstico, se generó un gráfico de *loading* (Figura 6), descrito por PC1 y PC2, el cual describe qué variables tienen mayor peso o *loadings*. Se consideró de *loading* mayor importancia o *high factor loading* a aquellas variables que tenían un valor de *loading* mayor a 0,25 considerando su valor absoluto (Peterson et al., 2000).

Las variables con *loading* positivo (figura 6) con respecto al eje de PC1 (hemoglobina, eritrocitos, hematocrito) tienen directa correlación con los usuarios de buen pronóstico, considerando que ambos se ubican en el eje positivo de PC1 en el gráfico de *scores* (figura 5).

Las variables con *loading* positivo con respecto al eje de PC2 (nitrogeno ureico, urea, creatinina, INR, PCR), tienen directa correlación con los usuarios de mal pronóstico, considerando que ambos se ubican en el eje positivo de PC2 en el gráfico *scores* (figura 5), representado por línea punteada de color rojo.



**Figura 6. Pesos o *loadings* de parámetros clínicos según PCA**

Gráfico de pesos o *loadings* proyectados en las nuevas dimensiones descritas por el componente principal 1 y 2. Se muestran las variables con valores de *loading*  $>|0.25|$ . En paréntesis está expresado el porcentaje de varianza de cada componente principal. Después de la tabla general

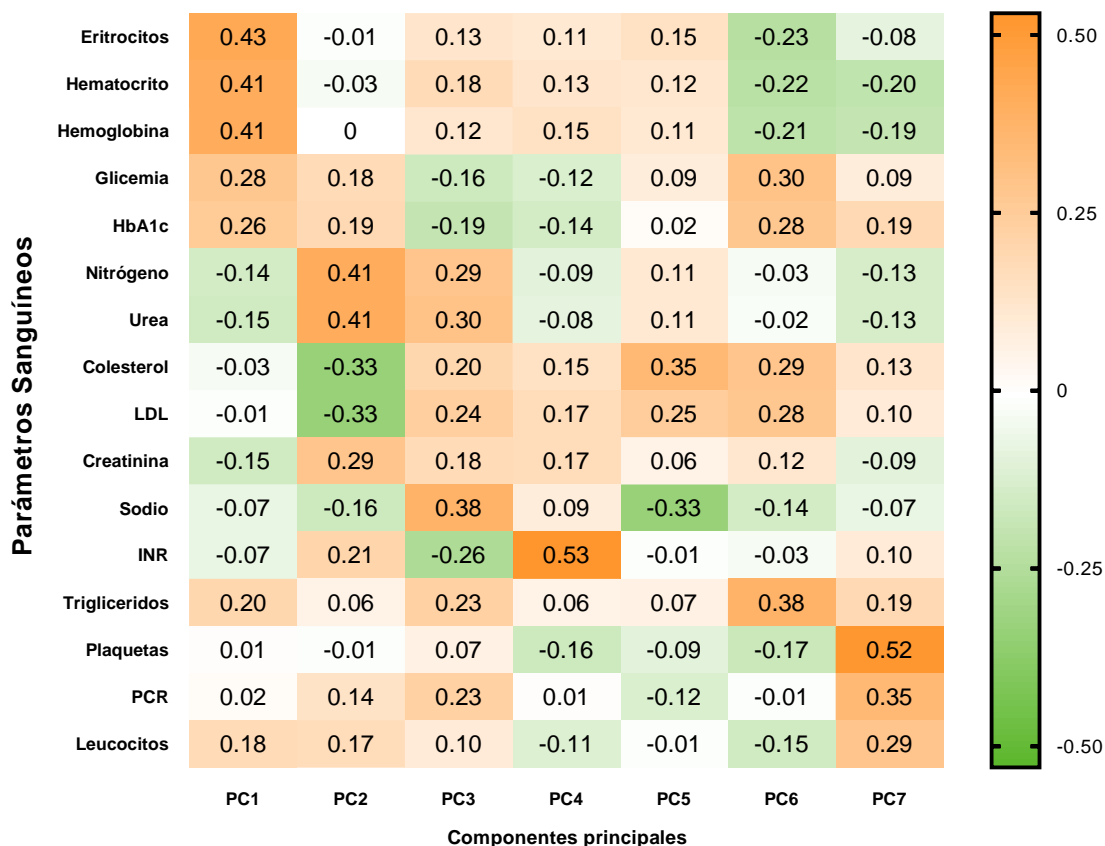
Se puede observar en la tabla 2, los valores de loading para cada parámetro clínico medido en los componentes principales analizados mediante PCA: Los valores positivos de loading desplazan la distribución de puntos o *scores* hacia el lado positivo del eje del componente principal, mientras que los valores negativos de loading desplazan la distribución de *scores* hacia el lado negativo del componente principal. Se detalla además en la última fila de cada columna, la varianza de cada componente principal y la varianza acumulada.

**Tabla 2. Loadings de 7 componentes principales según PCA**

|                      | <b>PC1</b> | <b>PC2</b> | <b>PC3</b> | <b>PC4</b> | <b>PC5</b> | <b>PC6</b> | <b>PC7</b> |
|----------------------|------------|------------|------------|------------|------------|------------|------------|
| Eritrocitos          | 0.43       | -0.01      | 0.13       | 0.11       | 0.15       | -0.23      | -0.08      |
| Hematocrito          | 0.41       | -0.03      | 0.18       | 0.13       | 0.12       | -0.22      | -0.20      |
| Hemoglobina          | 0.41       | 0.00       | 0.12       | 0.15       | 0.11       | -0.21      | -0.19      |
| Glicemia             | 0.28       | 0.18       | -0.16      | -0.12      | 0.09       | 0.30       | 0.09       |
| HbA1c                | 0.26       | 0.19       | -0.19      | -0.14      | 0.02       | 0.28       | 0.19       |
| Nitrógeno Ureico     | -0.14      | 0.41       | 0.29       | -0.09      | 0.11       | -0.03      | -0.13      |
| Urea                 | -0.15      | 0.41       | 0.30       | -0.08      | 0.11       | -0.02      | -0.13      |
| Colesterol total     | -0.03      | -0.33      | 0.20       | 0.15       | 0.35       | 0.29       | 0.13       |
| LDL                  | -0.01      | -0.33      | 0.24       | 0.17       | 0.25       | 0.28       | 0.10       |
| Creatinina           | -0.15      | 0.29       | 0.18       | 0.17       | 0.06       | 0.12       | -0.09      |
| Sodio                | -0.07      | -0.16      | 0.38       | 0.09       | -0.33      | -0.14      | -0.07      |
| INR                  | -0.07      | 0.21       | -0.26      | 0.53       | -0.01      | -0.03      | 0.10       |
| Triglicéridos        | 0.20       | 0.06       | 0.23       | 0.06       | 0.07       | 0.38       | 0.19       |
| Plaquetas            | 0.01       | -0.01      | 0.07       | -0.16      | -0.09      | -0.17      | 0.52       |
| PCR                  | 0.02       | 0.14       | 0.23       | 0.01       | -0.12      | -0.01      | 0.35       |
| Leucocitos           | 0.18       | 0.17       | 0.10       | -0.11      | -0.01      | -0.15      | 0.29       |
| Varianza %           | 13.94      | 12.67      | 8.59       | 8.03       | 7.71       | 6.77       | 6.13       |
| Varianza acumulada % | 13.94      | 26.61      | 35.2       | 43.23      | 50.94      | 57.71      | 63.84      |

Peso o *loadings* de los parámetros sanguíneos analizados en los siete primeros componentes principales en el PCA. Los valores  $>|0.25|$  se muestran en negrita (Peterson et al.,2000).

Con el objetivo de visualizar de una manera más ilustrativa los *loadings* de los parámetros sanguíneos analizados, se realizó un gráfico de calor o *heat map* (figura 7), en el cual la intensidad de colores está relacionada de modo directamente proporcional con el valor absoluto del *loading* en cada componente principal. Los componentes principales observados son vectores, los cuales podrían considerarse una variable nueva, constituida por las variables originales que aportan al componente principal en mayor o menor medida, según sus valores de *loading*. En el componente principal 1 (PC1), se observa que las variables de mayor importancia (*loading* >|0.25) y correlación positiva con dicho componente, corresponden a parámetros del perfil Hemograma (Hemoglobina;0.41, Hematocrito;0.41, Eritrocito;0.43), seguidos de parámetros asociados al metabolismo y regulación de glucosa (Glicemia;0.28 y Hemoglobina glicosilada;0.26). Al analizar el componente principal 2 (PC2), las variables de mayor importancia (*loading* >|0.25) y correlación negativa, pertenecen al perfil lipídico (LDL; -0.33 y Colesterol total; -0.33), y al analizar aquellas de mayor *loading* con correlación positiva, destacan las variables asociadas a la función renal, (Nitrógeno ureico;0.41, Urea;0.41, Creatinina;0.29).



**Figura 7.** Pesos o *Loadings* de los parámetros sanguíneos analizados en los 7 componentes principales según PCA. La gradiente de color indica la el peso y la dirección de la correlación entre los parámetros sanguíneos analizados y sus componentes principales. Los valores positivos están representados en naranja y los valores negativos están representados en verdes. La intensidad del color incrementa según aumenta el peso o *loading*.



## DISCUSIÓN Y CONCLUSIÓN

Las agrupaciones formadas espontáneamente según HCA, tal como se ilustra en el dendrograma (figura 1), revelan que existe concordancia entre la agrupación espontánea y los parámetros fisiológicos y/o fisiopatológicos. Este concepto es importante ya que aunque no aplicáramos una variables de clasificación, las variables clínicas se ordenan de manera coherente, esta distribución revela que la matriz de datos analizada es coherente y que es posible identificar diferencias y similitudes entre diferentes grupos de pronóstico. Si bien, aas agrupaciones ilustradas por los dendrogramas en los grupos de buen (figura 2) y mal (figura 3) pronóstico presentan similitudes entre sí, es posible observar ciertas diferencias puntuales en la cercanía de ciertos clusters e incluso el porcentaje de similitud de variación de comportamiento entre variables dentro de un mismo cluster. Esto implica que existen variaciones en los valores de los parámetros clínicos analizados según el grupo pronóstico. Cabe señalar que al analizar los pacientes por HCA (Datos no mostrados por la gran cantidad de pacientes), la distribución de los grupos es mezclada indicando que dentro de la matriz existen variable que contribuyen con ruido de dato a una buena clasificación

Considerando que las distribuciones observadas (figura 4) son espontáneas debido al carácter no supervisado de la herramienta utilizada, se puede indicar que PC1 y PC2 describen la separación con una progresión lógica y no forzada entre buen y mal pronóstico, alineado con la marca de clase categórica asignada según variable clasificatoria modified rankin scale (mRS).

Con respecto a los demás componentes principales analizados (PC3 a PC7), representan una menor varianza con respecto a PC1 y PC2 (tabla 2) y no presentan una distribución espontánea que ilustre la separación asignada por la variable de clase buen y mal pronóstico (figura 4).

Es importante señalar que las 3 variables sanguíneas del hemograma hematocrito, hemoglobina y eritrocitos, presentan valores de *loading* similares (figura 7), lo que responde a la relación de dependencia de las variables, demostrando que el análisis no supervisado es capaz de agrupar elementos que están estrechamente relacionados unos con otros. Este mismo fenómeno se observa en las variables nitrógeno ureico y urea. Con respecto a las variables identificadas como de mayor relevancia en el pronóstico del ACV isquémico, es posible observar que, de acuerdo con el componente principal 1 (PC1) (figura 6) en aquellos usuarios de buen pronóstico, existen mayores niveles en los parámetros asociados al hemograma, tales como hematocrito, hemoglobina y eritrocitos.

Si bien los estudios clínicos han relacionado niveles elevados de hematocrito y hemoglobina con una mayor incidencia de ACV (Yang et al., 2018) (Alonso et al., 2021), debemos señalar que en este caso se está analizando el fenómeno fisiopatológico post ACV isquémico. Si bien desde el punto de vista de la estadística clásica (tabla 1) los usuarios no presentan diferencias significativas en estos parámetros analizados, se puede observar una tendencia de mayores valores de hematocrito y hemoglobina en los pacientes con buen pronóstico. Se puede hipotetizar que los procesos de regeneración vascular, requieren elementos importantes derivados de una buena oxigenación del tejido dañado (Erning et al., 2020) y como consecuencia, la elevación de los eritrocitos circulantes, pueden llevar a una mejor oxigenación y por defecto a un mejor proceso de cicatrización celular y mejora en el pronóstico del usuario post ACV isquémico.

Con respecto al segundo componente principal (PC2), se puede observar que valores alterados del perfil renal, tales como la elevación de creatinina y urea, tienen directa correlación con un mal pronóstico. Diversos estudios (Pan et al., 2020) (Mosenzon et al., 2020) sugieren que existe una relación entre la función renal y su impacto en la recuperación de los usuarios post ACV isquémico. Además, diferente evidencia de estudios

experimentales (Zhao et al.,2020) y observacionales sugieren la existencia de una conversación o *cross-talk* entre el cerebro y el riñón luego del ACV isquémico. Esto podría, por una parte, identificar parámetros renales como la urea y la creatinina como factores de mal pronóstico de ACV isquémico, lo que concuerda con diversos estudios disponibles (Schrock et al., 2012) (RackIMuthu et al., 2020). Por otro lado, también podría anticipar el deterioro renal que existe luego del ACV isquémico, lo que significa un deterioro en la funcionalidad del usuario luego de este evento (Arnold et al.,2018). Es preciso señalar también, que la elevación de la urea de forma aguda, está vinculada a la destrucción de tejidos y proteínas, evento que ocurre en la muerte celular del tejido afectado por la falta de irrigación sanguínea durante el ACV isquémico (Cui et al.,2020).

Otras variables que también deben ser consideradas como factores de mal pronóstico según los análisis realizados para PC2 (figura 7) son la glicemia (loading factor;0.28) y hemoglobina glicosilada (Hb1Ac) (loading factor;0.26), lo que concuerda con la literatura científica disponible (Wada et al., 2018) (Lau et al.,2019).

Finalmente, se observa que bajos niveles de Lipoproteínas de alta densidad (LDL) y de colesterol total están asociados a un mejor pronóstico (loading factor;-0.33 para PC2), lo que coincide con la evidencia que sugiere que, mayores niveles de colesterol están relacionados a una mayor incidencia de ACV y peor pronóstico (Chen et al.,2018) (Deng et al.,2018). El estudio revela que los algoritmos de machine learning empleados permiten identificar factores de buen y mal pronóstico luego del ACV. Los factores identificados significan información relevante que permitirá mejorar futuros modelos predictivos desarrollados a nivel local, los que serán foco de futuro desarrollo.

Como futuro trabajo se realizará un ensayo Análisis Supervisado mediante la técnica de Mínimos cuadrados parciales (PLS), la cual en breve utiliza los vectores de los diferentes

componentes principales combinándolos mediante rotación de sus ejes. Esto generaría uno o varios algoritmos de predicción derivado de una regresión lineal con las variables categóricas de buen y mal comportamiento como variables dependientes de las variables clínicas medidas.

### REFERENCIAS BIBLIOGRÁFICAS:

1. Lavados, P. M., Hoffmeister, L., Moraga, A. M., Vejar, A., Vidal, C., Gajardo, C., ... & Sacks, C. (2021). Incidence, risk factors, prognosis, and health-related quality of life after stroke in a low-resource community in Chile (ÑANDU): a prospective population-based study. *The Lancet Global Health*.
2. Johnson, W., Onuma, O., Owolabi, M., & Sachdev, S. (2016). Stroke: a global response is needed. *Bulletin of the World Health Organization*, 94(9), 634.
3. Ministerio de Salud – Gobierno de Chile. 2017. Ataque Cerebrovascular. [online] Available at: <[https://www.minsal.cl/ataque\\_cerebral/](https://www.minsal.cl/ataque_cerebral/)> [Accessed 17 March 2020].
4. Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., ... & Bhutta, Z. A. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258), 1204-1222.
5. 2013. Guía Clínica Auge, Accidente Cerebro Vascular Isquémico En Personas De 15 Años O Más. 2nd ed. [ebook] Santiago de Chile: ministerio de salud, pp.47-92. Available at: <<https://www.minsal.cl/portal/url/item/7222754637e58646e04001011f014e64.pdf>> [Accessed 17 March 2020].
6. Amarenco, P., Denison, H., Evans, S. R., Himmelmann, A., James, S., Knutsson, M., ... & Reiner, P. (2021). Ticagrelor added to aspirin in acute ischemic stroke or transient ischemic attack in prevention of disabling stroke: a randomized clinical trial. *JAMA neurology*, 78(2), 177-185.
7. Westerlind, E., Hörsell, D., & Persson, H. C. (2020). Different predictors after stroke depending on functional dependency at discharge: a 5-year follow up study. *BMC neurology*, 20(1), 1-8.

8. Li, J., Muraki, I., Imano, H., Cui, R., Yamagishi, K., Umesawa, M., ... & Iso, H. (2020). Serum uric acid and risk of stroke and its types: the Circulatory Risk in Communities Study (CIRCS). *Hypertension Research*, 43(4), 313-321.
9. Rivera, R., Amudio, C., Brunetti, E., Catalan, P., Sordo, J. G., Badilla, L., ... & Nogueira, R. G. (2020). Breaking the breach in Latin America: A pilot study of mechanical thrombectomy in the public healthcare system in Chile. *Interventional Neuroradiology*, 1591019920956405.
10. Ministerio de Salud de Chile, (2020). Ataque cerebrovascular isquémico en personas de 15 años y más. <http://www.supersalud.gob.cl/difusion/665/w3-article-611.html>
11. Stinear, C. M., Lang, C. E., Zeiler, S., & Byblow, W. D. (2020). Advances and challenges in stroke rehabilitation. *The Lancet Neurology*, 19(4), 348-360.
12. Zuo, L., Zhang, L., Zu, J., Wang, Z., Han, B., Chen, B., ... & Yao, H. (2020). Circulating circular RNAs as biomarkers for the diagnosis and prediction of outcomes in acute ischemic stroke. *Stroke*, 51(1), 319-323.
13. Lavados, P. M., Sacks, C., Prina, L., Escobar, A., Tossi, C., Araya, F., ... & Álvarez, G. (2007). Incidence, case-fatality rate, and prognosis of ischaemic stroke subtypes in a predominantly Hispanic-Mestizo population in Iquique, Chile (PISCIS project): a community-based incidence study. *The Lancet Neurology*, 6(2), 140-148.
14. Sennfält, S., Pihlgård, M., Norrving, B., Ullberg, T., & Petersson, J. (2021). Ischemic stroke patients with prestroke dependency: Characteristics and long-term prognosis. *Acta Neurologica Scandinavica*, 143(1), 78-88.
15. Luna-Matos, M., Mcgrath, H., & Gaviria, M. (2007). Manifestaciones neuropsiquiátricas en accidentes cerebrovasculares. *Revista chilena de neuro-psiquiatría*, 45(2), 129-140.
16. Han, E., Kim, T. H., Koo, H., Yoo, J., Heo, J. H., & Nam, H. S. (2019). Heterogeneity in costs and prognosis for acute ischemic stroke treatment by comorbidities. *Journal of neurology*, 266(6), 1429-1438.
17. Yaghi, S., Herber, C., Boehme, A. K., Andrews, H., Willey, J. Z., Rostanski, S. K., ... & Boden-Albala, B. (2017). The association between diffusion MRI-defined infarct volume and NIHSS score in patients with minor acute stroke. *Journal of Neuroimaging*, 27(4), 388-391.
18. Abzhandadze, T., Reinholdsson, M., & Sunnerhagen, K. S. (2020). NIHSS is not enough for cognitive screening in acute stroke: a cross-sectional, retrospective study. *Scientific reports*, 10(1), 1-8.
19. Zhao, X. J., Li, Q. X., Liu, T. J., Wang, D. L., An, Y. C., Zhang, J., ... & Zheng, F. X. (2018). Predictive values of CSS and NIHSS in the prognosis of patients with acute cerebral infarction: a comparative analysis. *Medicine*, 97(39).
20. Abzhandadze, T., Reinholdsson, M., & Sunnerhagen, K. S. (2020). NIHSS is not enough for cognitive screening in acute stroke: a cross-sectional, retrospective study. *Scientific reports*, 10(1), 1-8.

21. Broderick, J. P., Adeoye, O., & Elm, J. (2017). Evolution of the modified Rankin scale and its use in future stroke trials. *Stroke*, 48(7), 2007-2012.
22. Chen, X., Li, J., Anderson, C. S., Lindley, R. I., Hackett, M. L., Robinson, T., ... & Enchanted Investigators. (2021). Validation of the simplified modified Rankin scale for stroke trials: experience from the ENCHANTED alteplase-dose arm. *International Journal of Stroke*, 16(2), 222-228.
23. Cramer, S. C., Le, V., Saver, J. L., Dodakian, L., See, J., Augsburger, R., ... & Janis, S. (2021). Intense arm rehabilitation therapy improves the modified Rankin Scale score: association between gains in impairment and function. *Neurology*, 96(14), e1812-e1822.
24. Lattanzi, S., Pulcini, A., Corradetti, T., Rinaldi, C., Zedde, M. L., Ciliberti, G., & Silvestrini, M. (2020). Prediction of outcome in embolic strokes of undetermined source. *Journal of Stroke and Cerebrovascular Diseases*, 29(1), 104486.
25. Alfieri, D. F., Lehmann, M. F., Flauzino, T., de Araújo, M. C. M., Pivoto, N., Tirolla, R. M., ... & Reiche, E. M. V. (2020). Immune-Inflammatory, Metabolic, Oxidative, and Nitrosative Stress Biomarkers Predict Acute Ischemic Stroke and Short-Term Outcome. *Neurotoxicity research*, 38, 330-343.
26. Lin, C. H., Hsu, K. C., Johnson, K. R., Fann, Y. C., Tsai, C. H., Sun, Y., ... & Taiwan Stroke Registry Investigators. (2020). Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry. *Computer Methods and Programs in Biomedicine*, 190, 105381.
27. Wiens, J., Price, W. N., & Sjoding, M. W. (2020). Diagnosing bias in data-driven algorithms for healthcare. *Nature medicine*, 26(1), 25-26.
28. Lanas, F., & Seron, P. (2021). Facing the stroke burden worldwide. *The Lancet Global Health*, 9(3), e235-e236.
29. Kogan, E., Twyman, K., Heap, J., Milentijevic, D., Lin, J. H., & Alberts, M. (2020). Assessing stroke severity using electronic health record data: a machine learning approach. *BMC medical informatics and decision making*, 20(1), 1-8.
30. Peng, C., Chen, Y., Kang, Z., Chen, C., & Cheng, Q. (2020). Robust principal component analysis: A factorization-based approach with linear complexity. *Information Sciences*, 513, 581-599.
31. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*.
32. Escudero, C., Acurio, J., López, E., Rodríguez, A., Benavente, A., Lara, E., & Korzeniewski, S. J. (2021). Vascular endothelial growth factor and poor prognosis after ischaemic stroke. *European Journal of Neurology*, 28(5), 1759-1764.
33. Wilson, J. L., Hareendran, A., Grant, M., Baird, T., Schulz, U. G., Muir, K. W., & Bone, I. (2002). Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. *Stroke*, 33(9), 2243-2246.

34. Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90-95.
35. Peterson, R. A. (2000). A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing letters*, 11(3), 261-275.
36. Yang, R., Wang, A., Ma, L., Su, Z., Chen, S., Wang, Y., ... & Wang, C. (2018). Hematocrit and the incidence of stroke: a prospective, population-based cohort study. *Therapeutics and clinical risk management*, 14, 2081.
37. Alonso, A., Aparicio, F. H. J., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, F. A. P., ... & Stokes, F. A. (2021). Heart Disease and Stroke Statistics—2021 Update. *Circulation*, 2021(143), e00-e00.
38. Erning, K., & Segura, T. (2020). Materials to promote recovery after stroke. *Current Opinion in Biomedical Engineering*.
39. Pan, X., Zhou, F., Shen, R., Zhu, Y., Arima, H., Yang, J., & Zhou, J. (2020). Influence of renal function on stroke outcome after mechanical thrombectomy: a prospective cohort study. *BMC neurology*, 20, 1-6.
40. Mosenzon, O., Bain, S. C., Heerspink, H. J., Idorn, T., Mann, J. F., Persson, F., ... & LEADER Trial Investigators. (2020). Cardiovascular and renal outcomes by baseline albuminuria status and renal function: Results from the LEADER randomized trial. *Diabetes, Obesity and Metabolism*, 22(11), 2077-2088.
41. Zhao, Q., Yan, T., Chopp, M., Venkat, P., & Chen, J. (2020). Brain–kidney interaction: renal dysfunction following ischemic stroke. *Journal of Cerebral Blood Flow & Metabolism*, 40(2), 246-262.
42. Schrock, J. W., Glasenapp, M., & Drogell, K. (2012). Elevated blood urea nitrogen/creatinine ratio is associated with poor outcome in patients with ischemic stroke. *Clinical neurology and neurosurgery*, 114(7), 881-884.
43. RacklMuthu, S., IShwaRa, P. R. P., & SuReSh, S. (2020). Complications Associated with Osmotic Therapy in Acute Stroke Patients-A Prospective Longitudinal Observational Study. *Journal of Clinical & Diagnostic Research*, 14(7).
44. Arnold, J., Ng, K. P., Sims, D., Gill, P., Cockwell, P., & Ferro, C. (2018). Incidence and impact on outcomes of acute kidney injury after a stroke: a systematic review and meta-analysis. *BMC nephrology*, 19(1), 1-10.
45. Cui, X., Wei, W., Qin, X., Hou, F., Zhu, J., & Li, W. (2020). Comparison of routine hematological indicators of liver and kidney function, blood count and lipid profile in healthy people and stroke patients. *PeerJ*, 8, e8545.
46. Wada, S., Yoshimura, S., Inoue, M., Matsuki, T., Arihiro, S., Koga, M., ... & Toyoda, K. (2018). Outcome prediction in acute stroke patients by continuous glucose monitoring. *Journal of the American Heart Association*, 7(8), e008744.

47. Lau, L. H., Lew, J., Borschmann, K., Thijs, V., & Ekinçi, E. I. (2019). Prevalence of diabetes and its effects on stroke outcomes: A meta-analysis and literature review. *Journal of diabetes investigation*, 10(3), 780-792.
48. Chen, Y. W., Li, C. H., Yang, C. D., Liu, C. H., Chen, C. H., Sheu, J. J., ... & Taiwan Stroke Registry Investigators. (2017). Low cholesterol level associated with severity and outcome of spontaneous intracerebral hemorrhage: Results from Taiwan Stroke Registry. *PloS one*, 12(4), e0171379.
49. Deng, Q. W., Li, S., Wang, H., Lei, L., Zhang, H. Q., Gu, Z. T., ... & Yan, F. L. (2018). The short-term prognostic value of the triglyceride-to-high-density lipoprotein cholesterol ratio in acute ischemic stroke. *Aging and disease*, 9(3), 498.

## ANEXOS

### Anexo 1: Respuesta Comité Ético Científico HCHM





ORD N° 31.-

ANT: Solicitud de Investigación N° 20/20

MAT: Respuesta del CEC

CHILLAN, 14 de julio de 2020

A: Antonia Belén Benavente Rubio:

Me permito comunicar a usted, que el Comité Ético Científico (CEC) del Hospital Clínico Herminia Martín, integrado por el Dr. Carlos Escudero (presidente); Abg. Claudia Cabrera; Mg. Marcela Espinoza (secretaria); Dr. Patricio Oliva; Dra. Ninette Poseck; Dr. Nain Hormazábal; Q.F. Jessica Fernández y Sr. René Riveros; teniendo en cuenta los siguientes documentos:

1. Carta de presentación de proyecto dirigida al Director del Hospital.
2. Carta compromiso de la jefatura correspondiente.
3. Protocolo de solicitud de investigación científica
4. Currículo actualizado del investigador principal
5. Currículo actualizado de cada uno de los investigadores
6. Y documentos anexos de respaldo solicitados por el CEC (encuestas, carta Gantt, cuestionario resuelto de las "preguntas al investigador")

Luego de la lectura y análisis de los mencionados documentos, habiéndose excluido de ello el Dr. Escudero por conflicto de interés al ser parte del equipo de investigadores, y en consideración de los criterios relevantes del protocolo presentado, que incluyen: validez científica; relación riesgo-beneficio favorable para el sujeto de estudio; detallado proceso de protección de la confidencialidad; viabilidad de la investigación en el centro propuesto; investigador idóneo para llevar a cabo el proyecto; utilidad social y novedad científica, reconociendo la relevancia social y científica del estudio; la relación riesgo-beneficio favorable analizada para el estudio; los resguardos de la confidencialidad que se explicitan y fueron solicitados por el CEC para el manejo de la información; y la naturaleza retrospectiva del estudio que no involucra necesariamente la intervención al sujeto de estudio, decidió entregar una dispensa del Consentimiento Informado".

COMITÉ ÉTICO CIENTÍFICO <http://hospitaleschilano.cl/web/comite-etico-cientifico> [www.hospitaleschilano.cl](http://www.hospitaleschilano.cl)  
[etica.sites@redbasi.cl](mailto:etica.sites@redbasi.cl) Fono 427433 Red Mtral 587431

## ANEXO 2: Tabla de operativización de variables

| <b>Variable</b>       | <b>Definición</b>  | <b>Parámetro de análisis</b>   |
|-----------------------|--|--|
| Pronóstico            | Grado de incapacidad o dependencia en las actividades diarias post ACV según Modified Rankin scale (mRS)   | Se clasificará como:<br>-Buen pronóstico mRS $\leq 2$<br>-Mal pronóstico mRS $\geq 3$  |
| Edad                  | Es el tiempo transcurrido entre el día, mes y año de nacimiento y el día mes y año en que ocurrió y/o se registró el hecho vital.  | Años cumplidos registrados en base de datos al momento de ingresar al centro hospitalario.   |
| Sexo                  | Es la característica biológica que permite clasificar a los seres humanos en hombres y mujeres.  | Se clasificará como:<br>-Hombre<br>-Mujer  |
| Antecedentes mórbidos | Afección o enfermedad de una duración de 3 meses o más, que no representa un cuadro agudo y que permanece de manera definitiva en el usuario.                                    | Se clasificará como:<br>-Hipertensión Arterial<br>-Diabetes Mellitus<br>-Epilepsia<br>-Asma<br>-Dislipidemia<br>-Otros   |
| Hemograma             | Descripción y número de las diversas clases de células que se encuentran en una cantidad determinada de sangre y de las proporciones entre ellas.                                | Se medirá según valor sérico:<br>Eritrocitos: millones/mm <sup>3</sup><br>Hemoglobina: g/dL<br>Hematocrito: %<br>Plaquetas: miles/mm <sup>3</sup><br>Leucocitos: miles/mm <sup>3</sup> |
| Perfil lipídico       | Descripción del estado del metabolismo de los lípidos corporales.  | Se medirá según valor sérico:<br>Colesterol total: mg/dL<br>HDL: mg/dL<br>LDL: mg/dL<br>Triglicéridos: mg/dL   |
| Perfil hepático       | Análisis de sangre en el que se mide la presencia de algunas enzimas, proteínas y bilirrubina en sangre, con el objetivo de determinar si existe alguna alteración en el hígado. | Albúmina: g/dL<br>Fosfatasa alcalina: UI/L<br>Transaminasa GGT: UI/L<br>Bilirrubina: mg/dL<br>Tiempo protombina: segundos  |

|                          |  |   |
|--------------------------|--|---|
| Perfil renal             | Análisis empleado para evaluar la función renal.   | Creatinina: mg/dL<br>Urea: mg/dL<br>Nitrógeno ureico: mg/dL |
| Electrolitos plasmáticos | Análisis sanguíneo que mide la cantidad presente de sodio, cloro y potasio.              | Cloro: mEq/L<br>Sodio: mEq/L<br>Potasio: mEq/L              |
| Hemoglobina glicosilada  | Cantidad de glucosa adherida a la hemoglobina, promedio glicémico de últimos 3 meses.    | HbA1c: %  |
| Glicemia                 | Medida de concentración de glucosa libre en la sangre.                                   | Glicemia: mg/dL   |
| Proteína c reactiva      | Proteína plasmática circulante, que aumenta sus niveles en respuesta a la inflamación.   | PCR: mg/l   |
| Pruebas de coagulación   | Análisis de sangre que comprueban cómo funciona la coagulación sanguínea y sus factores. | INR: segundos<br>TTPA: segundos<br>TP: segundos             |

### Anexo 3: Escala de Rankin modificada (mRS)

|          |   |   |
|----------|---|---|
| <b>0</b> | <b>Sin síntomas</b>                     |   |
| <b>1</b> | <b>Sin incapacidad importante</b>       | Capaz de realizar sus actividades y obligaciones habituales   |
| <b>2</b> | <b>Incapacidad leve</b>                 | Incapacidad de realizar algunas de sus actividades previas, pero capaz de velar por sus intereses y asuntos sin ayuda   |
| <b>3</b> | <b>Incapacidad moderada</b>             | Síntomas que restringen significativamente su estilo e vida o impiden su subsistencia totalmente autónoma (p. ej. necesitando alguna ayuda)                                     |
| <b>4</b> | <b>Incapacidad moderadamente severa</b> | Síntomas que impiden claramente su subsistencia independiente aunque sin necesidad de atención continua (p. ej. incapaz para atender sus necesidades personales sin asistencia) |
| <b>5</b> | <b>Incapacidad severa</b>               | Totalmente dependiente, necesitando asistencia constante día y noche  |
| <b>6</b> | <b>Muerte</b>                           |   |