

UNIVERSIDAD DE CHILE  
FACULTAD DE MEDICINA  
ESCUELA DE POSTGRADO  
PROGRAMA DE GRADOS ACADÉMICOS



Programa de Magíster en Informática Médica  
**“Identificación precoz de *outliers* superiores según  
análisis de Grupos relacionado por Diagnóstico (GRD)”**  
TESIS PARA OPTAR EL GRADO DE MAGISTER EN INFORMÁTICA MÉDICA

**Alumno(a):** Jorge Carlos Montoya Zumaeta

**Director(a) de Tesis:** Prof. Eduardo Tobar Almonacid

**Co-director de Tesis:** Prof. Héctor Ramírez Cabrera

**Co-guía de Tesis:** Prof. Gloria Henríquez Díaz

.....

**Firma Director(a) de Tesis**

Prof. Eduardo Tobar Almonacid

.....

**Firma Co-Director(a) de Tesis**

Prof. Héctor Ramírez Cabrera

.....

**Firma Co-Guía de Tesis**

Prof. Gloria Henríquez Díaz

.....

**Firma de alumno**

Jorge Carlos Montoya Zumaeta

**UNIVERSIDAD DE CHILE  
FACULTAD DE MEDICINA  
ESCUELA DE POSTGRADO  
PROGRAMA DE GRADOS ACADÉMICOS**



**“IDENTIFICACIÓN PRECOZ DE *OUTLIERS* SUPERIORES  
SEGÚN ANÁLISIS DE GRUPOS RELACIONADO POR  
DIAGNÓSTICO (GRD)”**

**JORGE CARLOS MONTOYA ZUMAETA**

**TESIS PARA OPTAR EL GRADO DE MAGISTER EN INFORMÁTICA MÉDICA**

<b>Director de Tesis:</b>	Prof. Eduardo Tobar Almonacid
<b>Co-director de Tesis:</b>	Prof. Héctor Ramírez Cabrera
<b>Co-guía de Tesis:</b>	Prof. Gloria Henríquez Díaz

**2025**

## **Agradecimientos**

A mi familia, quienes han sido un pilar fundamental en mi vida. El apoyo incondicional, ánimo constante y la frase “lo que empiezas debes terminarlo” fueron esenciales para sostenerme a lo largo de este proceso académico y personal.

Al Prof. Dr. Eduardo Tobar, Director General del HCUCH, por su orientación académica y su guía, a pesar de las múltiples responsabilidades asociadas a su cargo.

Agradezco especialmente al Prof. Dr. Héctor Ramírez, a la Prof. Dra. Gloria Henríquez y al Prof. Dr. Víctor Riquelme, del Centro de Modelamiento Matemático (CMM), por su apoyo en los aspectos más desafiantes del proyecto, especialmente en el ámbito del modelamiento matemático, área que ha representado un reto significativo para mí. Asimismo, expreso mi reconocimiento al Proyecto Basal CMM (FB210005) por su constante respaldo, acompañamiento y valiosa contribución al desarrollo de este trabajo de investigación.

A mis amigos, colegas intensivistas y jefes técnicos de la Unidad de Paciente Crítico (UPC) del Hospital Clínico Universidad de Chile (HCUCH), quienes siempre demostraron comprensión y apoyo. De manera especial, agradezco a mi jefe Dra. Nivia Estuardo, por su permanente estímulo, por facilitar la compatibilización de mis responsabilidades asistenciales con las demandas del magíster y por brindarme el tiempo necesario para avanzar y terminar la tesis.

A mis amiga y compañera del programa Estefanía por su presencia constante, su preocupación genuina y su valioso respaldo durante las etapas más exigentes de la presentación de la tesis.

## **TABLA DE CONTENIDO**

1.	RESUMEN.....	6
2.	ABSTRACT .....	8
3.	INTRODUCCIÓN.....	10
3.1.	Antecedentes .....	10
3.2.	Marco teórico .....	15
3.3.	Problema.....	27
4.	HIPÓTESIS.....	28
5.	OBJETIVOS.....	29
5.1.	Objetivo General .....	29
5.2.	Objetivos Específicos.....	29
6.	METODOLOGÍA.....	30
6.1.	Diseño del estudio .....	30
6.2.	Población y muestra .....	30
6.3.	Fuentes de información y variables.....	30
6.4.	Calidad de datos.....	31
6.5.	Definición operacional de outlier superior:.....	33
6.6.	Preprocesamiento de datos.....	33
6.7.	Modelos predictivos.....	34
6.8.	Aspectos éticos y manejo de datos: .....	36
6.9.	Software y entorno.....	37
7.	RESULTADOS.....	38
7.1.	Comprender el proceso de clasificación GRD del HCUCH (OE1) .....	38
7.2.	Describir la base de datos GRD y caracterizar sus variables, identificando patrones asociados a estancia hospitalaria prolongada (OE2). .....	41
7.2.1.	Descripción base de datos GRD y DAU del HCUCH.....	41

7.2.2.	<i>Evaluación de la calidad de datos.....</i>	41
7.2.3.	<i>Preprocesamiento y análisis de datos .....</i>	44
7.2.4.	<i>Transformaciones de variables.....</i>	45
7.2.5.	<i>Análisis descriptivo y univariado .....</i>	47
7.3.	<i>Identificar factores de riesgo asociados a outliers superiores mediante análisis conjunto de la base de datos GRD y RCE del HCUCH (OE3). ....</i>	49
7.3.1.	<i>Análisis de los outliers superiores e inliers.....</i>	49
7.3.2.	<i>Análisis de los GRD con mayor outliers superiores.....</i>	50
7.3.3.	<i>Factores de riesgo asociados a outliers superiores .....</i>	53
7.4	<i>Comparar modelos predictivos basados en algoritmos de aprendizaje automático que permita anticipar, desde la atención de urgencia, a los pacientes con mayor riesgo de transformarse en outliers superiores (OE4). ....</i>	58
7.5	<i>Identificar y analizar los factores que dificultan y limitan la capacidad predictiva de los modelos, comparados para la predicción de outliers superiores (OE5).....</i>	63
8.	<i>DISCUSIÓN.....</i>	65
9.	<i>CONCLUSIONES.....</i>	69
10.	<i>PERSPECTIVAS FUTURAS.....</i>	71
11.	<i>BIBLIOGRAFÍA.....</i>	73
12.	<i>ANEXOS .....</i>	78

## 1. RESUMEN

El sistema de Grupos Relacionados por Diagnóstico (GRD) constituye una herramienta estandarizada utilizada internacionalmente para clasificar los episodios hospitalarios a partir de datos administrativos y clínicos. Fue concebido originalmente para clasificación retrospectiva, destinada a describir los episodios hospitalarios una vez finalizada la atención. El propósito de esta herramienta es clasificar y organizar los egresos hospitalarios de manera efectiva, agrupándolos según patrones comparables de diagnósticos, procedimientos, edad y comorbilidades. Esto permite una visión estructurada y detallada de la casuística de los pacientes.

Cuando el GRD se utiliza con fines de evaluación de desempeño o eficiencia hospitalaria, surgen casos que exceden los parámetros estadísticos definidos para su grupo: los denominados *outliers* superiores, que es definido por aquel paciente cuya duración de hospitalización excede significativamente el tiempo esperado establecido para su respectivo GRD. Estos no representan errores del sistema de codificación, sino variaciones extremas dentro de un grupo clínico homogéneo, que reflejan la complejidad real del proceso asistencial, la presencia de complicaciones, o diferencias en la organización del cuidado.

El problema central radica en que la asignación del GRD ocurre de manera retrospectiva (al alta médica del paciente) por lo que no permite anticipar qué casos evolucionarán hacia ser *outliers* superiores. Así, aunque el GRD describe adecuadamente lo ocurrido, no ofrece herramientas para intervenir de forma temprana. Esta brecha entre descripción y predicción constituye el punto de partida de este estudio, orientado a desarrollar modelos que permitan identificar precozmente a los pacientes con riesgo de ser *outliers* superiores al momento del ingreso hospitalario.

El presente estudio tiene como objetivo identificar de manera temprana a los pacientes con riesgo de convertirse en *outliers* superiores al momento de su ingreso por el Servicio de Urgencia del Hospital Clínico de la Universidad de Chile (HCUCH), utilizando modelos predictivos basados en algoritmos de aprendizaje automático (*machine learning*). Para ello, se empleó la base de datos GRD del HCUCH, que comprende 13.760 egresos hospitalarios correspondientes a los años 2018, 2019 y 2022. Se incluyeron pacientes de 15 años o más cuyo ingreso se produjo a través del servicio de urgencias.

A continuación, y acorde a los objetivos específicos que se señalarán, se describe el proceso de clasificación GRD, que integra información demográfica, administrativa y clínica codificada según los sistemas CIE-10 (diagnósticos) y CIE-9 (procedimientos). Este procedimiento permite asignar un GRD específico a cada hospitalización, constituyendo la base para el análisis comparativo de estancias.

Posteriormente, se realiza un análisis descriptivo de la cohorte seleccionada ( $\geq 15$  años, ingreso por urgencia), identificándose 568 códigos GRD únicos. Los pacientes clasificados como *outliers* superiores correspondieron al 4,6 % del total de egresos ( $n = 632$ ), con una mediana de 27 días de hospitalización, significativamente mayor que la de los *inliers* (6 días;  $p < 0,001$ ). Las variables asociadas a una mayor probabilidad de ser *outlier* superior fueron el ingreso a unidades críticas, el triaje de alta prioridad y diagnósticos iniciales como sepsis, insuficiencia respiratoria o abscesos periamigdalinos.

A continuación, se evalúan diferentes modelos predictivos supervisados obteniéndose un AUC ROC de 0,653 para Regresión Logística y 0,663 para Naive Bayes, el rendimiento general fue moderado, condicionado por el desbalance de clases y la heterogeneidad diagnóstica.

Entre las principales limitaciones se identificaron: la baja prevalencia de *outliers*, la similitud entre variables clínicas iniciales de ambos grupos y la influencia de factores no observables al ingreso hospitalario. Estos resultados destacan la necesidad de fortalecer los modelos predictivos mediante la incorporación de bases de datos ampliadas, variables clínicas dinámicas y características longitudinales del proceso de atención.

En conclusión, este trabajo sienta las bases para el desarrollo de un modelo predictivo precoz capaz de detectar pacientes con riesgo de ser *outliers* o *inliers* hospitalarios, integrando variables clínicas iniciales disponibles al ingreso. En el futuro, este modelo podría complementarse con herramientas de inteligencia artificial, orientadas a optimizar la asignación de recursos y fortalecer la calidad asistencial

## **2. ABSTRACT**

The Diagnosis-Related Group (DRG) system is an internationally recognized, standardized tool used to classify hospital episodes based on administrative and clinical data. It was originally conceived for retrospective classification, intended to describe hospital episodes once care had concluded. The purpose of this tool is to effectively classify and organize hospital discharges, grouping them according to comparable patterns of diagnoses, procedures, age, and comorbidities. This allows for a structured and detailed view of patient casuistry.

When the DRG is used for hospital performance or efficiency evaluation, cases that exceed the defined statistical parameters for their group emerge: these are called high-end outliers. A high-end outlier is defined as a patient whose length of stay significantly exceeds the expected time established for their respective DRG. These do not represent coding errors, but rather extreme variations within a clinically homogeneous group, reflecting the real complexity of the care process, the presence of complications, or differences in the organization of care.

The central problem lies in the fact that DRG assignment occurs retrospectively (upon patient discharge), which does not allow for anticipating which cases will evolve into high-end outliers. Thus, while the DRG adequately describes what happened, it does not offer tools for early intervention. This gap between description and prediction is the starting point of this study, which is focused on developing models that allow for the early identification of patients at risk of becoming high-end outliers at the time of hospital admission.

The present study aims to identify, at the moment of admission through the Emergency Department (ED) of the Hospital Clínico de la Universidad de Chile (HCUCH), those patients at risk of becoming high-end outliers, using predictive models based on machine learning algorithms. For this purpose, the HCUCH DRG database, comprising 13,760 hospital discharges corresponding to the years 2018, 2019, and 2022, was utilized. Patients aged 15 years or older whose admission occurred through the emergency service were included.



In line with the specific objectives outlined below, the DRG classification process is described, which integrates demographic, administrative, and clinical information coded according to the ICD-10 (diagnoses) and ICD-9 (procedures) systems. This procedure allows a specific DRG to be assigned to each hospitalization, forming the basis for the comparative analysis of lengths of stay.

Subsequently, a descriptive analysis of the selected cohort (age >15 years, ED admission) was performed, identifying 568 unique DRG codes. Patients classified as high-end outliers accounted for 4.6% of total discharges ( $n = 632$ ), with a median length of stay of 27 days, significantly longer than that of inliers (6 days;  $p < 0.001$ ). Variables associated with a higher probability of being a high-end outlier included admission to critical care units, high-priority triage, and initial diagnoses such as sepsis, respiratory failure, or peritonsillar abscess.

Next, different supervised predictive models were evaluated, yielding an AUC ROC of 0.653 for Logistic Regression and 0.663 for Naive Bayes; the overall performance was moderate, conditioned by class imbalance and diagnostic heterogeneity.

Key limitations identified were: the low prevalence of outliers, the similarity between initial clinical variables of both groups, and the influence of factors not observable at hospital admission. These results highlight the need to strengthen predictive models by incorporating expanded databases, dynamic clinical variables, and longitudinal characteristics of the care process.

In conclusion, this work lays the foundation for developing an early predictive model capable of detecting patients at risk of being hospital outliers or inliers, integrating initial clinical variables available at admission. In the future, this model could be complemented with artificial intelligence tools aimed at optimizing resource allocation and strengthening care quality.

### **3. INTRODUCCIÓN**

#### **3.1. Antecedentes**

Los grupos relacionados a diagnóstico (GRD) son un sistema de clasificación de pacientes que son clínicamente consistentes y que tienen patrones similares de utilización de recursos medidos a través de la duración de la estadía hospitalaria (1). La clasificación GRD tiene fundamentalmente dos componentes, la primera es la lógica de agrupación que clasifican decenas de miles de servicios de hospitalización en un número limitado de grupos en función de la similitud tanto de diagnósticos como de tratamientos que recibieron los pacientes y la segunda componente son los recursos relevantes que se utilizaron durante la hospitalización (2). La clasificación GRD de la hospitalización de un paciente es única y excluyente, es decir que sólo puede asignarse un sólo código GRD y los demás códigos son descartados.

Fue originado en Estados Unidos por Fetter et al en la década de 1960, siendo la primera generación de sistemas de GRD desarrollada en 1976 (1). En Chile, el año 2002 en el Hospital Clínico de la Universidad Católica fue presentado un proyecto FONDEF para la implementación de sistema de GRD y en noviembre del año 2009 se lleva a cabo el proyecto de “Implementación de Sistema de Grupos Relacionados al Diagnóstico en Hospitales de Alta Complejidad” con 16 instituciones participantes a nivel nacional. La incorporación de más hospitales fue progresiva, el año 2011 se integran 23 hospitales y en el año 2012 otros 18 y en el año 2020, se incorporan 65 establecimientos de la Red Pública (55 de alta complejidad y 10 de mediana complejidad) (3). La Unidad de GRD del HCUCH se creó a partir de la firma del convenio entre el Banco Interamericano de Desarrollo (BID) y la Universidad de Chile para el desarrollo del Proyecto Instrumentos para el Mejoramiento de la Gestión y Productividad en el Mercado Hospitalario Chileno en el año 2007 (4).

La producción de un hospital es un proceso difícil de medir debido a lo complejo de las actividades asistenciales hospitalarias y a las dificultades inherentes a cualquier actividad de servicios. La clasificación de pacientes según metodología GRD fue diseñado para gestionar los costos de atención médica y mantener operaciones sostenibles para los pacientes hospitalizados y se ha convertido en un componente importante de los pagos de atención médica en muchos países para promover el riesgo compartido entre las aseguradoras y los prestadores (clínicas u hospitales) (5). En el año 2020 ha sido incorporado como principal

mecanismo de pago de FONASA a los hospitales públicos, siendo una herramienta valiosa para el seguimiento de costos y asignación de recursos. La codificación normalmente se realiza de forma retrospectiva después del alta con el egreso hospitalario (3).

Para que un egreso hospitalario sea clasificado a un GRD determinado, se debe contar con información de la hospitalización, esto es conocido como conjunto mínimo básico de datos (CMBD) que constituye un resumen o extracto de información administrativa y clínica estandarizada, recogida a partir del egreso hospitalario (6, 7). Cada uno de los componentes del CMBD serán las variables necesarias como fuente de información para el proceso de agrupación a través de una herramienta informática, existiendo un analizador clínico estadístico disponible en cada uno de los centros, se logra entregar informes para la gestión clínica y toma de decisiones documentada (3).

Los algoritmos de agrupación de GRD proceden de los mismos agrupadores base, como los HCFA-GRDs (Health Care Finance Administration DRGs) creados en 1983, los AP-GRDs (All Patient DRGs) de 1988, los APR-GRDs (All Patient Refined DRGs) de 1994 y los IRDRGs (International Refined DRGs) de 2000. Las diferentes terminologías surgen por la evolución de los agrupadores y los objetivos que se buscan cubrir. Muchos países han adaptado versiones de los GRD a sus necesidades específicas de información y prácticas clínicas locales (8). En Chile el agrupador que se utiliza es el International Refined GRD (IR-GRD) versión 3.0, que es usado tanto en el ámbito público como privado y su particularidad es que permite la codificación de pacientes provenientes como egresos de la actividad ambulatoria y de la actividad con hospitalizados. Además, emplea un eje de agrupación centrado en el procedimiento, pero no cualquiera, sino el que consuma mayor recurso o que esté más asociado con la categoría diagnóstica mayor (CDM)(3). EL IR-GRD v 3.0 tiene 1081 GRD totales, de los cuales 792 (73%) pertenecen a GRD hospitalizados, tiene 3 niveles de severidad y 3 riesgos de mortalidad (9). Los pacientes de un GRD individual (grupo) muestran un curso similar de la enfermedad y una duración comparable de la estancia media hospitalaria, por lo tanto, los gastos también deben ser comparables, lo que permite vincular un grupo GRD específico con el reembolso a la institución sanitaria (10).

Los IR-GRDs tienen 6 dígitos, cada posición nos proporciona la siguiente información (3) (Fig N° 01):

- Dígitos 1 y 2 corresponden al código de la categoría diagnóstica mayor (CDM), que va desde el 01 al 23 (Anexo N° 01), por ejemplo, la categoría CDM 07 pertenece a: “Enfermedades y trastornos de hígado, sistema biliar y páncreas”
- Dígito 3 corresponde al tipo de GRD, que va desde el 0 al 9 (Anexo N° 02), y se interpreta como el tipo de episodio, el 4 pertenece a médico hospitalizado.
- Dígitos 4 y 5 corresponde al GRD específico, va del 00-99, es único para CDM y tipo de GRD, por ejemplo, según 07 (CDM) y 4 (médico hospitalizado 4) tenemos según IR-GRD 3.0:
  - o 074 - 10 MH Hepatitis alcohólica y cirrosis
  - o 074 - 11 MH Neoplasias de páncreas y sistema hepatobiliar
  - o 074 - 12 MH Enfermedades de páncreas excepto neoplasias
  - o 074 - 13 MH Enfermedades hepáticas excepto neoplasias, cirrosis o hepatitis alcohólica
  - o 074 - 14 MH Otras enfermedades del sistema biliar
- Dígito 6 corresponde al nivel de severidad, en el caso de ambulatorios el número es cero y en el caso de hospitalizados tiene 3 niveles de severidad: 1 = menor, 2 = moderada y 3 = mayor.
- Existe un dígito 7 que no está explícito en el código GRD, pero representa un indicador de riesgo de mortalidad: 1 = menor, 2 = moderada y 3 = mayor.
- Tanto la severidad como el riesgo de mortalidad es determinado por las comorbilidades, edad y diagnósticos secundarios que presenta el paciente.
- También existen códigos de error o inagrupables (que empiezan con el número 99) que son por inconsistencia entre diagnósticos, codificación, tipo de GRD o datos administrativos. Por ejemplo asignar un procedimiento como “cesárea” a sexo “hombre”.

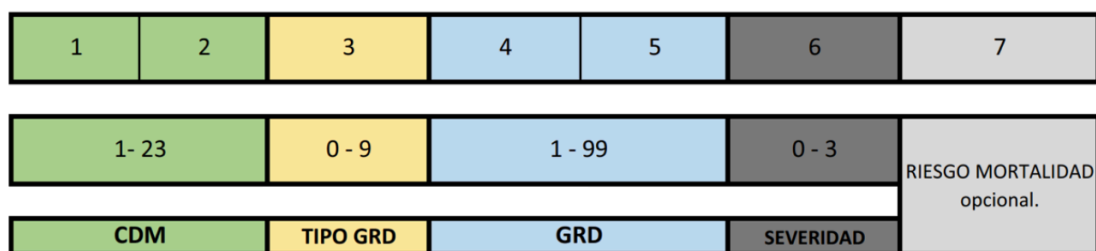


Figura N° 01 Lógica de numeración del IR-GRD (3)

Con los datos clínicos ingresados al agrupador se pueden obtener indicadores de producción hospitalaria, como la frecuencia en la casuística hospitalaria, los IR-GRD más frecuentes, la información de estancia media hospitalaria y peso medio para ese grupo específico; adicionalmente también se pueden obtener indicadores de eficiencia como el porcentaje de *outliers* superiores, estancia media (EM), estancia media de la Norma (EMN), estancia media ajustada por el funcionamiento (EMAF) e índice funcional (IF) (4) . Esto permite conocer y cuantificar lo que el hospital produce y estos datos tienen relevancia para la gestión clínica y financiera ya que nos permite evaluar el desempeño de diferentes servicios clínicos a través del tiempo y con otras instituciones de salud. Actualmente el estándar nacional está dado por la Norma MINSAL 2018-2019 (9).

Los ***outliers* superiores** son los pacientes cuya estancia fue mayor del punto de corte superior para ese GRD en específico (fórmula:  $\text{Percentil } 75 + 1,5 \times [\text{Percentil } 75 - \text{Percentil } 25]$ ) para la norma MINSAL 2018-2019 vigente a la fecha (4,11). Su porcentaje constituye un indicador relevante para detectar problemas en la gestión hospitalaria, ya que estos casos suelen analizarse retrospectivamente y pueden reflejar complicaciones clínicas, o ineficiencias en la atención, es importante porque implican un consumo de recursos desproporcionado en relación con su GRD asignado.

Con respecto a los factores de riesgo para estancias hospitalarias prolongadas según GRD (*outliers* superiores), se han identificado diversas variables, las cuales varían según la población y el tipo de hospital. Un estudio retrospectivo en un hospital universitario traumatológico en Suiza investigó los factores predictivos de estancias prolongadas, encontrando que estaban asociados a la edad (65 años o más), múltiples diagnósticos, comorbilidades, infecciones y complicaciones perioperatorias (12). En Turquía, un estudio

en 15 hospitales docentes mostró que los *outliers* superiores representaban solo el 4.4% de los pacientes, pero acumulaban el 24.5% de los días de hospitalización, con factores influyentes como quemaduras, uso de drogas/alcohol y casos neonatales (13). En Portugal, se identificó que los *outliers* superiores constituían el 3.9% de los casos, representando el 19.2% de los días de hospitalización. Los factores más influyentes fueron la edad avanzada, el ingreso de emergencia, comorbilidades y el tipo de hospital, especialmente en grandes hospitales universitarios (14).

El sistema GRD es una herramienta de gestión hospitalaria y financiera para abordar calidad de atención, eficiencia hospitalaria y problemas de costo, pero la pregunta es si reflejan adecuadamente la complejidad de los diferentes grupos de pacientes hospitalizados (2), como por ejemplo, en accidente cerebro vascular (15), trauma (16), parto (17), entre otros. Básicamente, en el sistema GRD se han buscado productos homogéneos que consuman recursos hospitalarios de forma similar, por lo que es una excelente herramienta de gestión (18).

Una de las principales preocupaciones con respecto a la implementación de sistema GRD es su efecto en la calidad de atención, como por ejemplo favorecer las altas precoces. En un metaanálisis de 29 estudios en China se evaluó la efectividad del sistema GRD en la calidad de la atención hospitalaria, el cual demostró que fue efectivo en reducir la duración de la estancia hospitalaria, sin embargo, no se encontró un efecto significativo en indicadores de calidad como las tasas de reingreso dentro de los 30 días ni en la mortalidad hospitalaria (19). En Suiza la implementación de GRD en una unidad de cuidados intensivos (UCI) no afectó la política de admisión, excepto por un aumento de pacientes internos con una baja gravedad clínica de la enfermedad, y tampoco afectó la mortalidad ni la estancia hospitalaria (10).

Existe un modelo basado en procesamiento de lenguaje natural (PLN) DRG-LLaMA que fue introducido para predecir código GRD a partir del egreso hospitalario del conjunto de datos del MIMIC IV, el modelo mejoró la precisión de la asignación del GRD, alcanzado un 52% de exactitud, superando otros modelos, aunque este modelo se basa en datos post alta hospitalaria, nos indica que el uso de evoluciones clínicas tempranas nos podrían permitir predecir GRD (20).

Existen estudios que han intentado predecir en forma precoz el código GRD mediante un modelos de PLN, con el objetivo de anticipar costos y optimizar la asignación de recursos en tiempo real. En uno de ellos, los datos de entrada fueron las evoluciones clínicas dentro de las primeras 24 a 48 horas de ingreso a una unidad de cuidados intensivos (UCI), el modelo fue de aprendizaje profundo (*Deep learning*) basado en *convolutional neural networks* (CNN), resultando en una predicción bastante precisa tanto para el sistema MS-DRG y APR-DRG (21). Otro estudio utilizó como modelos CNN y *gated recurrent unit* (GRU), siendo este último el que tuvo mayor precisión en los diagnósticos primarios (22).

### 3.2. Marco teórico

#### **Preprocesamiento de datos:**

En el ámbito del análisis secundario de registros clínicos electrónicos (RCE), el proceso de preprocesamiento de datos se establece como una fase metodológica indispensable y crítica para la validez de la investigación. Dada la naturaleza de los RCE, cuyo propósito original es la asistencia clínica y no la investigación, los datos inherentes suelen presentar desafíos significativos como la heterogeneidad en los formatos, el ruido, los errores de registro y una prevalencia notable de valores faltantes. Como se detalla en la literatura especializada (MIT Critical Data, 2016), esta etapa abarca una serie de acciones rigurosas que inician con la selección precisa de la cohorte y la limpieza de datos, seguidas por la integración y transformación de variables clínicas. Un componente esencial es la gestión de la información incompleta, donde se aplican diversas estrategias de imputación para mitigar el sesgo potencial. La finalidad última del preprocesamiento es la obtención de un conjunto de datos estructurado, "ordenado" (*tidy*) y confiable, capaz de soportar la aplicación de modelos estadísticos avanzados y de *machine learning* para generar inferencias causales y predictivas robustas (23).

Dentro de las principales técnicas se incluyen (24):

1. *Data Cleaning* (Limpieza de datos): aborda los principales problemas de calidad presentes en los datos clínicos y administrativos, tales como valores faltantes, registros incompletos, ruido, duplicidades e inconsistencias. Existen diversas estrategias para manejar valores

faltantes, que incluyen desde la omisión de registros cuando la ausencia es extensa, hasta métodos de imputación como el uso de una constante global, la media del atributo, la media dentro de cada clase o el valor más probable mediante modelos estadísticos. Asimismo, existen técnicas para reducir el ruido en los datos, como *binning*, regresión y *clustering*, que permiten suavizar variaciones anómalas y reconocer *outliers* que podrían distorsionar el análisis. Esta etapa es fundamental para garantizar un conjunto de datos limpio, coherente y confiable antes de aplicar modelos predictivos.

2. Data Integration (Integración de datos): Combina información proveniente de múltiples fuentes, resolviendo inconsistencias estructurales y semánticas entre ellas. La integración requiere identificar y eliminar redundancias, armonizar formatos de variables y unificar esquemas de datos, lo que resulta especialmente relevante en entornos hospitalarios donde conviven sistemas administrativos, clínicos y de codificación. Además, incluye la agregación de datos en estructuras multidimensionales, como *data cubes*, que permiten resumir información por períodos, servicios o grupos diagnósticos. Una integración adecuada mejora la consistencia global del *dataset* y facilita la construcción de una base consolidada para análisis descriptivos y predictivos.

3. Data Transformation (Transformación de datos): Esta etapa adapta los datos a un formato adecuado para su análisis estadístico o su incorporación en algoritmos de *machine learning*. Existen técnicas como la normalización, que ajusta la escala de las variables para evitar que atributos con rangos mayores dominen el modelo; la agregación, utilizada para sintetizar valores repetidos o generar indicadores globales; y la generalización, que permite convertir valores muy específicos en categorías más manejables mediante jerarquías conceptuales. También se incluyen métodos de suavizado para reducir variabilidad no deseada y facilitar la detección de patrones. La transformación convierte los datos crudos en representaciones más estables, comparables y aptas para el modelado predictivo.

4. Data Reduction (Reducción de datos): Busca disminuir el volumen de información manteniendo la esencia del comportamiento de los datos. Incluye la reducción de dimensionalidad que permiten compactar múltiples atributos correlacionados en componentes más informativos. También abarca métodos de reducción de numerosidad como muestreo, histogramas y *clustering*, los cuales representan grandes volúmenes de datos



mediante resúmenes o grupos homogéneos. Asimismo, la discretización, por *binning*, histogramas o métodos basados en entropía, permite transformar variables continuas en intervalos más adecuados para ciertos algoritmos. Estas técnicas mejoran la eficiencia computacional, reducen el ruido y fortalecen la capacidad de generalización de los modelos predictivos.

### **Machine learning (ML):**

Es una rama de la inteligencia artificial que desarrolla algoritmos capaces de aprender patrones a partir de datos y mejorar su desempeño sin ser programados explícitamente para cada situación. En el ámbito de la salud, estos métodos permiten construir modelos predictivos aplicados a tareas como estratificación de riesgo, detección temprana de eventos clínicos y optimización de procesos hospitalarios.

En el aprendizaje supervisado, los modelos utilizan datos etiquetados (por ejemplo, pacientes clasificados como *outlier* superior o no *outlier*) para aprender una función que prediga dicho resultado en nuevos casos. Este enfoque es el más utilizado en salud porque permite modelar relaciones complejas entre múltiples variables clínicas, demográficas y administrativas, utilizando algoritmos como regresión logística, Random Forest o XGBoost. Su objetivo principal es generalizar adecuadamente a datos no vistos, manteniendo precisión y estabilidad en contextos clínicos reales (25).

Por otro lado, el aprendizaje no supervisado trabaja con datos sin etiquetas y busca descubrir estructuras ocultas, como agrupamientos o patrones latentes. En medicina, se utiliza para identificar perfiles de pacientes, patrones de uso hospitalario o subgrupos clínicamente significativos, contribuyendo a una mejor comprensión de la heterogeneidad en las poblaciones de estudio. Estos métodos complementan al aprendizaje supervisado al aportar una caracterización preliminar que puede guiar o enriquecer los modelos predictivos.

Ambos paradigmas comparten desafíos clave en aplicaciones biomédicas, como la necesidad de manejar datos escasos o ruidosos, garantizar interpretabilidad y respetar la coherencia clínica. Los modelos predictivos en salud deben no solo alcanzar buen rendimiento, sino también demostrar robustez, capacidad de generalización y seguridad, especialmente en decisiones sensibles o contextos asistenciales complejos (26). De esta forma, el ML se

posiciona como una herramienta fundamental para apoyar la toma de decisiones clínicas y mejorar la eficiencia en los sistemas hospitalarios.

Finalmente, el uso creciente de ML en medicina se apoya en la disponibilidad de registros clínicos electrónicos, bases administrativas y grandes repositorios biomédicos. En este entorno, los modelos predictivos se han vuelto una herramienta estratégica para optimizar decisiones clínicas, priorizar recursos, anticipar complicaciones y mejorar la eficiencia hospitalaria. No obstante, su utilidad depende de la calidad de los datos, de una validación adecuada y de su integración ética y responsable dentro de los flujos asistenciales.

Para el presente estudio se emplearon algoritmos representativos de diversas familias de modelos, con el fin de comparar su desempeño predictivo bajo condiciones de desbalance de clases. A continuación, se presenta el fundamento conceptual de cada uno.

- Regresión Logística (RegLog): La regresión logística es un modelo estadístico ampliamente utilizado para la predicción de variables dependientes binarias. Su fundamento radica en la estimación de la probabilidad de un evento mediante una función logística que transforma una combinación lineal de predictores en valores entre 0 y 1. Su interpretabilidad y robustez frente a diversos tipos de datos la convierten en una herramienta común en el ámbito de la epidemiología y la predicción clínica (27).
  - Principales ventajas: interpretabilidad, eficiencia computacional.
  - Limitaciones: dificultad para capturar relaciones no lineales sin transformaciones adicionales.
  
- Random Forest: es un método de ensamblaje (ensemble) basado en múltiples árboles de decisión construidos sobre subconjuntos aleatorios de datos y predictores. Su principio central es la reducción de la varianza mediante el promedio de múltiples modelos débiles, lo que mejora la estabilidad y generalización. Es especialmente útil para modelar relaciones no lineales y detectar interacciones entre variables (28).
  - Ventajas: robustez, manejo de datos faltantes, no requiere supuestos de distribución.
  - Limitaciones: menor interpretabilidad en comparación con modelos lineales.

- XGBoost (Extreme Gradient Boosting): es un algoritmo basado en *boosting* de gradiente, que construye modelos secuenciales en los que cada árbol corrige los errores del anterior. Se caracteriza por su alta eficiencia computacional, regularización incorporada y excelente rendimiento en conjuntos de datos estructurados. Es uno de los algoritmos más usados en competencias de predicción por su capacidad para manejar desbalance y capturar patrones complejos (29).
  - Ventajas: regularización L1/L2, manejo eficiente de grandes volúmenes de datos.
  - Limitaciones: puede sobreajustarse si no se controla el número de árboles.
  
- LightGBM: es un algoritmo de *gradient boosting* optimizado desarrollado por Microsoft. Utiliza técnicas como *leaf-wise growth* y *histogram-based splitting*, lo que permite un entrenamiento significativamente más rápido que otros métodos de *boosting*. Es especialmente eficiente en conjuntos de datos con miles o millones de registros (30).
  - Ventajas: alta velocidad, bajo consumo de memoria, manejo nativo de datos categóricos.
  - Limitaciones: riesgo de sobreajuste en datos pequeños.
  
- Support Vector Classifier (SVC): es un algoritmo basado en la identificación del hiperplano que maximiza el margen entre clases en un espacio multidimensional. Permite la utilización de funciones núcleo (*kernels*) para modelar relaciones no lineales al proyectar los datos en espacios de mayor dimensión (31).
  - Ventajas: excelente desempeño en problemas de clasificación con fronteras complejas.
  - Limitaciones: alta demanda computacional en conjuntos de datos grandes y sensibilidad a la selección de hiperparámetros.
  
- Naive Bayes: es un clasificador probabilístico basado en el teorema de Bayes, bajo el supuesto de independencia condicional entre los predictores. A pesar de esta simplificación, es notablemente eficaz y presenta un excelente rendimiento en tareas con alta dimensionalidad o datos categóricos (25).
  - Ventajas: simple, rápido y eficiente incluso con conjuntos grandes.
  - Limitaciones: el supuesto de independencia puede no cumplirse en escenarios clínicos complejos.

- Red Neuronal Multicapa (MLP): pertenecen a la familia de modelos de aprendizaje profundo. Están compuestas por capas de neuronas artificiales conectadas entre sí que permiten aprender patrones no lineales y relaciones complejas entre variables. Se entrenan mediante retropropagación del error (*backpropagation*) y optimización por gradiente (25).
  - o Ventajas: alta capacidad de modelar patrones complejos.
  - o Limitaciones: requieren mayor cantidad de datos y ajustes finos de hiperparámetros para evitar sobreajuste.

### **Modelos predictivos**

Los modelos predictivos constituyen herramientas fundamentales para anticipar desenlaces clínicos y apoyar la toma de decisiones basadas en datos. En el ámbito sanitario, su utilidad ha crecido de manera exponencial debido a la disponibilidad de registros clínicos electrónicos, bases administrativas y métodos de aprendizaje automático capaces de manejar múltiples variables simultáneamente (30). Estos modelos permiten estimar riesgos individuales, identificar pacientes con mayor probabilidad de presentar determinados eventos y optimizar la planificación clínica y hospitalaria.

El desarrollo de un modelo predictivo clínico requiere seguir un marco metodológico riguroso que incluye la definición del objetivo, selección del desenlace, caracterización de la población, manejo de predictores, tratamiento de datos faltantes y validación interna y externa (32). Dicho proceso evita errores frecuentes como el sobreajuste, la selección no justificada de variables o la ausencia de evaluación de calibración, los cuales pueden comprometer la aplicabilidad real del modelo.

En salud, los modelos basados en *machine learning* han mostrado ventajas relevantes al capturar relaciones no lineales y patrones complejos. Algoritmos como Random Forest, XGBoost, LightGBM y Redes Neuronales se destacan por su capacidad para integrar grandes volúmenes de datos clínicos y mejorar el rendimiento predictivo en comparación con modelos tradicionales (33). Estas técnicas se complementan con modelos clásicos como la Regresión Logística y Naive Bayes, que siguen siendo ampliamente utilizados por su interpretabilidad y robustez (34). La literatura enfatiza que los modelos predictivos aplicados

en medicina deben ser evaluados no solo en términos de discriminación (AUC-ROC, sensibilidad, especificidad), sino también en calibración y utilidad clínica, elementos clave para su incorporación en decisiones asistenciales (32). Asimismo, deben someterse a validación en poblaciones independientes, considerando que los cambios epidemiológicos, organizacionales o tecnológicos pueden afectar la estabilidad del rendimiento.

Finalmente, la integración de modelos predictivos en salud representa una oportunidad para mejorar la eficiencia de los sistemas sanitarios, apoyar el diagnóstico temprano, reducir riesgos y distribuir recursos de forma más inteligente. Sin embargo, su utilidad depende estrictamente de la calidad de los datos, el rigor metodológico del desarrollo del modelo y el adecuado alineamiento con el contexto clínico de aplicación (35).

### **Pasos para el desarrollo de un modelo predictivo**

El proceso de construcción de un modelo predictivo en el ámbito de la salud requiere seguir una secuencia estructurada de etapas metodológicas para garantizar validez interna, generalización y utilidad clínica. De acuerdo con la guía de Efthimiou et al. publicada en The BMJ (2024), el desarrollo de un modelo de predicción clínica implica trece pasos fundamentales, cuya aplicación rigurosa permite evitar sesgos, mejorar la reproducibilidad y maximizar su impacto asistencial (32).

En primer lugar, es indispensable definir claramente el objetivo del modelo, incluyendo el desenlace a predecir, la población objetivo, el escenario clínico de aplicación y los usuarios que emplearán las predicciones. Esta definición inicial orienta todas las decisiones metodológicas posteriores. A continuación, se debe evaluar la evidencia existente, identificando modelos previos y sus limitaciones, lo que permite decidir entre desarrollar un modelo nuevo o actualizar uno ya existente.

Posteriormente, se procede a especificar los predictores candidatos, seleccionados con base en literatura biomédica, plausibilidad clínica y disponibilidad en práctica real. De igual importancia es la obtención y exploración de los datos, lo que incluye la detección de errores, revisión de distribuciones, análisis del patrón de datos faltantes y verificación de la calidad de mediciones. En paralelo, deben considerarse criterios de tamaño muestral, ya que un

número insuficiente de eventos o predictores incrementa el riesgo de sobreajuste (*overfitting*) y limita la capacidad de generalización.

La etapa subsecuente consiste en el manejo de los datos faltantes, recomendándose el uso de técnicas de imputación múltiple para conservar información y evitar sesgos asociados al análisis de casos completos. Una vez preparado el conjunto de datos, se procede a ajustar diferentes estrategias de modelamiento, que pueden incluir regresión logística, métodos penalizados o algoritmos de *machine learning*, según el tipo de datos y la complejidad del problema.

El desempeño del modelo debe ser evaluado mediante validación interna, utilizando métodos robustos como *bootstrapping* o *k-fold cross-validation*, con el fin de estimar honestamente la discriminación, calibración y desempeño global. Asimismo, cuando los datos lo permiten, es recomendable realizar validación interna-externa, que evalúa la estabilidad del modelo en distintos subgrupos o centros. Tras comparar distintas alternativas, se selecciona el modelo final considerando rendimiento, estabilidad e interpretabilidad.

Posteriormente, se recomienda realizar un análisis de curva de decisión para estimar la utilidad clínica del modelo a diferentes umbrales de riesgo, integrando beneficios y potenciales daños.

En conjunto, estos pasos aseguran que un modelo predictivo no solo presente adecuado rendimiento estadístico, sino también aplicabilidad clínica, transparencia metodológica y potencial de impacto en la toma de decisiones en salud.

### **Técnicas de sobremuestreo**

Las técnicas de sobremuestreo constituyen un conjunto de métodos destinados a corregir el desbalance en conjuntos de datos donde la clase minoritaria es escasamente representada. Esta problemática es frecuente en investigación clínica, donde los eventos relevantes (complicaciones, mortalidad, estancias prolongadas) suelen presentarse en baja proporción. El objetivo del sobremuestreo es aumentar la presencia de la clase minoritaria para mejorar

la capacidad de los modelos predictivos de identificar estos casos sin afectar el equilibrio global del *dataset*. En esencia, estas técnicas buscan reducir el sesgo hacia la clase mayoritaria y mejorar la sensibilidad en escenarios donde los casos minoritarios son clínicamente relevantes (36).

Entre las técnicas de sobremuestreo más utilizadas destaca SMOTE (*Synthetic Minority Oversampling Technique*), que genera ejemplos sintéticos mediante interpolación entre instancias reales de la clase minoritaria. Este procedimiento evita la simple duplicación de registros y permite que el espacio de representación de la clase minoritaria se expanda de manera más realista. Diversas variantes, como SMOTEENN, combinan el sobremuestreo con métodos de limpieza de ruido, produciendo *datasets* más depurados y modelos más estables. El estudio comparativo entre SMOTE y SMOTEENN demostró que esta última técnica ofrece un desempeño superior en métricas de exactitud, estabilidad y generalización, especialmente en contextos clínicos donde existen valores extremos y ruido en los datos (37).

Otra técnica de sobremuestreo utilizada para abordar el desbalance de clases es el método ADASYN (*Adaptive Synthetic Sampling Approach for Imbalanced Learning*) que constituye un aporte relevante al generar nuevos ejemplos sintéticos de manera adaptativa según la dificultad de aprendizaje de cada instancia minoritaria. A diferencia de métodos tradicionales como SMOTE, que asignan un número uniforme de muestras sintéticas a todos los casos minoritarios, ADASYN concentra la generación de datos en aquellos ejemplos ubicados en regiones más complejas del espacio de características, identificadas por la proporción de vecinos pertenecientes a la clase mayoritaria. Este enfoque permite reducir el sesgo hacia la clase dominante, mejorar la capacidad del modelo para aprender patrones difíciles y desplazar la frontera de decisión hacia zonas críticas del *dataset*. Los resultados experimentales muestran que ADASYN ofrece un desempeño superior en métricas sensibles al desbalance, mejorando el reconocimiento de la clase minoritaria sin deteriorar significativamente la mayoría (38).

A pesar del uso extendido de las técnicas de sobremuestreo, la evidencia empírica reciente ha mostrado que su efectividad depende del tipo de problema y del algoritmo utilizado. Un análisis exhaustivo aplicado a más de 1.500 modelos predictivos en bases de datos de salud

reveló que el *sobresampling* aleatorio no mejora el rendimiento de modelos clínicos en términos de discriminación (AUROC) ni en validación externa. De hecho, estas técnicas pueden introducir descalibración, afectando la estimación correcta del riesgo clínico. Aunque dicho sesgo puede corregirse mediante recalibración posterior, el estudio concluye que el sobremuestreo simple no aporta beneficios sustanciales en escenarios de datos masivos y heterogéneos como los registros clínicos electrónicos (36).

### Evaluación de modelos predictivos

La evaluación del desempeño de los modelos predictivos se realiza considerando de manera integral los componentes de discriminación, calibración y validación. La discriminación del modelo es analizada por Área Bajo la Curva ROC (AUC), que refleja su capacidad para diferenciar adecuadamente entre pacientes que presentan o no el evento de interés. La calibración, por su parte, se evalúa a través de curvas de calibración, el *calibration slope* y el *Brier Score*, permitiendo determinar la concordancia entre los riesgos predichos y los realmente observados. Asimismo, se implementan métodos de validación interna, como *k-fold cross-validation* y *bootstrapping*, con el fin de corregir el optimismo inherente al ajuste del modelo en la misma cohorte de entrenamiento. De manera complementaria, se considera la importancia de la validación externa, necesaria para estimar la transportabilidad del modelo hacia nuevas poblaciones, así como su utilidad clínica mediante *decision curve analysis*, evaluando el beneficio neto de su aplicación en escenarios reales de toma de decisiones (32).

Las métricas tradicionales de evaluación de modelos de clasificación, como *Accuracy*, *Precision*, *Recall* y *F1-score*, se basan en la asignación de una etiqueta final obtenida a partir de un umbral de decisión. A continuación, se describen brevemente las principales métricas utilizadas para evaluar el desempeño de los modelos:

1. **Exactitud (*Accuracy*):** proporción de predicciones correctas (positivas y negativas) respecto del total de casos evaluados.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde:



TP = Total Positivos

TN = Total Negativos

FP = Falsos Positivos

FN = Falsos Negativos

2. **Sensibilidad o Recall:** capacidad del modelo para identificar correctamente los casos positivos (*outliers*).

$$\text{Recall} = \frac{TP}{TP + FN}$$

Donde:

TP = Total Positivos

FN = Falsos Negativos

3. **Precisión:** proporción de verdaderos positivos entre todos los casos clasificados como positivos.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Donde:

TP = Total Positivos

FP = Falsos Positivos

4. **F1-Score:** media armónica entre precisión y sensibilidad; resume el equilibrio entre ambas.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. **AUC-ROC (Área bajo la curva ROC):** mide la capacidad global del modelo para discriminar entre clases (*outlier* vs *inlier*). Un valor de 0,5 indica rendimiento aleatorio y valores cercanos a 1,0 reflejan excelente discriminación.

Estas métricas se derivan de la matriz de confusión, que clasifica las predicciones en verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Su interpretación conjunta permite valorar la utilidad clínica de un modelo predictivo, especialmente en contextos con baja prevalencia del evento, como en este estudio. Estas métricas son ampliamente utilizadas debido a su simplicidad y su capacidad para ofrecer una primera aproximación al desempeño del modelo. Sin embargo, su valor interpretativo puede ser limitado en contextos donde existen clases desbalanceadas o cuando el modelo produce probabilidades más ricas que una simple predicción binaria (39).

### **Interpretabilidad de los datos**

La interpretabilidad en modelos de clasificación constituye un componente esencial en el desarrollo de sistemas de soporte clínico basados en inteligencia artificial. Aunque los modelos de aprendizaje automático han demostrado una capacidad creciente para procesar grandes volúmenes de datos y detectar patrones complejos, muchos de ellos operan como cajas negras, dificultando la comprensión de las relaciones entre variables y la lógica interna que conduce a una predicción específica. En entornos clínicos, donde las decisiones requieren justificación y trazabilidad, esta opacidad metodológica limita su adopción, ya que la ausencia de explicaciones claras reduce la confianza de los profesionales y plantea desafíos éticos y regulatorios (40,41).

Los clasificadores interpretables, como los árboles de decisión, las listas de reglas y los modelos basados en patrones de alta utilidad, buscan equilibrar rendimiento predictivo y transparencia. Estos modelos permiten comprender explícitamente cómo contribuye cada variable al proceso de clasificación, facilitando auditorías clínicas, detección de sesgos y la justificación de decisiones individuales. Estudios recientes han demostrado que enfoques híbridos, que combinan clasificadores complejos con módulos explicativos, pueden maximizar la fidelidad y minimizar la complejidad, generando explicaciones comprensibles sin sacrificar precisión. Métodos como LIME, SHAP y las listas de reglas optimizadas han adquirido relevancia al ofrecer explicaciones locales y globales aplicables incluso a modelos profundamente no lineales (42).

En el contexto sanitario, la interpretabilidad adquiere un valor adicional: permite integrar el razonamiento clínico humano con las salidas del modelo algorítmico. Esta sinergia es fundamental cuando se trabaja con decisiones sensibles como la predicción de estancias

prolongadas u *outliers* superiores, donde múltiples dimensiones clínicas, comorbilidades, gravedad inicial, variabilidad diagnóstica, interactúan de forma compleja y frecuentemente no observable en la admisión. En consecuencia, la interpretabilidad no solo mejora la utilidad del modelo, sino que fortalece la seguridad, reproducibilidad y aceptabilidad institucional, asegurando que los modelos predictivos actúen como herramientas complementarias y no sustitutos del juicio clínico.

### **3.3. Problema**

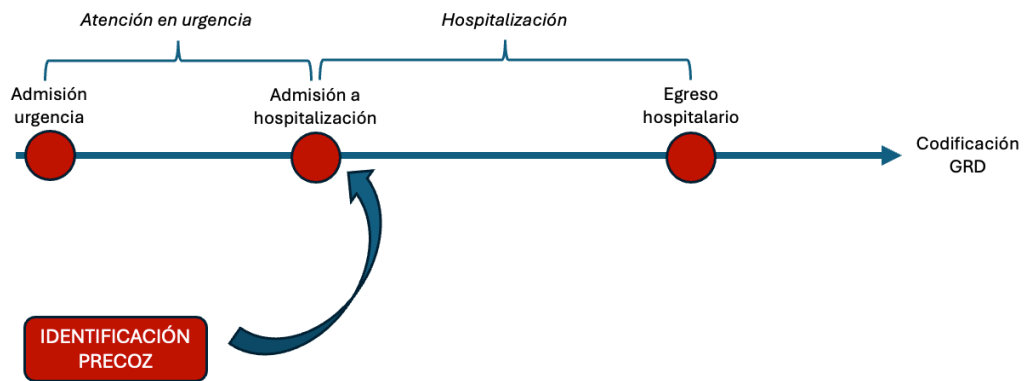
El porcentaje de *outliers* superiores constituye un indicador relevante para la identificación de ineficiencias o dificultades en el proceso de atención hospitalaria, ya que estos casos, definidos por estancias significativamente mayores a las esperadas según su GRD, suelen asociarse a complicaciones clínicas, retrasos diagnósticos o terapéuticos, fallas en la coordinación interservicios o limitaciones operativas de la institución. No obstante, el sistema GRD presenta una limitación estructural: su carácter retrospectivo. La clasificación y validación del GRD se realiza únicamente al momento del egreso hospitalario, de modo que la identificación de un paciente como *outlier* superior ocurre cuando la prolongación de la estancia ya se ha producido, dificultando una intervención temprana y reduciendo las oportunidades para modificar favorablemente su evolución.

En este contexto, la detección precoz de pacientes con riesgo de convertirse en *outliers* superiores adquiere un valor estratégico para la gestión clínica y administrativa. Anticipar esta condición desde el ingreso permitiría monitorear de manera continua el curso de la hospitalización, reconocer desviaciones tempranas en el proceso asistencial y activar intervenciones oportunas orientadas a optimizar la atención. Asimismo, el desarrollo de herramientas predictivas basadas en datos clínicos y administrativos disponibles desde el ingreso podría disminuir la ocurrencia de estancias excesivamente prolongadas, mejorar la eficiencia institucional y contribuir a un modelo de gestión hospitalaria más dinámico, preventivo y centrado en el paciente.

#### 4. HIPÓTESIS

Se plantea como hipótesis que el uso precoz de información de la base de datos GRD del HCUCH con la información disponible en el registro clínico electrónico (RCE) los Datos de Atención en Urgencia (DAU), permitiría anticipar la identificación de pacientes con riesgo de convertirse en *outliers* superiores. Este enfoque propone una detección precoz, entendida como la capacidad de predecir dicha condición antes de la admisión hospitalaria (Fig N° 02), es decir, utilizando exclusivamente los antecedentes clínicos, administrativos y demográficos obtenidos durante el proceso de atención en urgencia.

Figura N° 02 Flujo asistencia del paciente que ingresa a urgencia



## 5. OBJETIVOS

### 5.1. Objetivo General

OG. Identificar precozmente los *outliers* superiores mediante el análisis de la base de datos GRD y RCE del HCUCH.

### 5.2. Objetivos Específicos

OE1. Comprender el proceso de clasificación GRD del HCUCH.

OE2. Describir la base de datos GRD y caracterizar sus variables demográficas, clínicas y administrativas, identificando patrones asociados a estancia hospitalaria prolongada.

OE3. Identificar factores de riesgo asociados a *outliers* superiores mediante análisis conjunto de la base de datos GRD y RCE del HCUCH.

OE4. Comparar modelos predictivos basados en algoritmos de aprendizaje automático que permitan anticipar, desde la atención de urgencia, a los pacientes con mayor riesgo de transformarse en *outliers* superiores.

OE5. Identificar y analizar los factores que dificultan y limitan la capacidad predictiva de los modelos predictivos comparados para la predicción de *outliers* superiores.

## 6. METODOLOGÍA

### 6.1. Diseño del estudio

Estudio observacional, retrospectivo y analítico, basado en el uso de las bases de datos institucionales del HCUCH (GRD y DAU). Se llevó a cabo un análisis descriptivo y predictivo de los egresos hospitalarios correspondientes a los años 2018, 2019 y 2022, con el propósito de desarrollar un modelo capaz de identificar de manera temprana a los pacientes con riesgo de convertirse en *outliers* superiores

### 6.2. Población y muestra

- Población: se incluyeron todos los egresos hospitalarios del Hospital Clínico de la Universidad de Chile (HCUCH) codificados mediante GRD, cuya admisión se efectuó a través del Servicio de Urgencia.
- Criterios de inclusión: pacientes de 15 años o más que ingresaron por el Servicio de Urgencia y contaron con codificación GRD asociada al episodio hospitalario.
- Criterios de exclusión: se excluyeron los egresos correspondientes a los años 2020 y 2021, así como hospitalizaciones obstétricas, pediátricas, programadas o aquellas sin codificación GRD válida.
- Tamaño muestral: la cohorte final estuvo constituida por 13.760 egresos hospitalarios, correspondientes a los años 2018, 2019 y 2022.

### 6.3. Fuentes de información y variables

- Fuentes de información: se utilizaron la base de datos GRD junto con los DAU correspondientes a los años 2018 al 2022, de los registros clínicos electrónicos institucionales del sistema RCE TiCares.
- Variables: El examen de su estructura interna, obtenido mediante la función `df.info()`, evidencia la presencia de 27 variables de tipo entero (*int64*), 35 variables numéricas de punto flotante (*float64*) y 91 variables de tipo objeto (*object*). Estas últimas corresponden principalmente a datos textuales, categorías codificadas y campos derivados de registros clínicos o administrativos, incluyendo posibles formatos de fecha. Las variables seleccionadas fueron 15 y están agrupadas en tres dominios principales demográficas, administrativas y clínicas (Tabla N° 01):

Tabla N° 01: Clasificación de variables

DATOS	VARIABLES DE INTERÉS
Demográficos	Sexo
	Edad
	Nacionalidad
	Comuna
Administrativos	Servicio que ingresa desde la urgencia
	Tipo de previsión
Clínicos	Triage
	Presión arterial sistólica (PAS)
	Presión arterial diastólica (PAD)
	Presión arterial media (PAM)
	Frecuencia cardíaca (FC)
	Saturación de oxígeno (SatO2)
	Frecuencia respiratoria (FR)
	Escala de Glasgow
	Diagnóstico de urgencia CIE 10

#### 6.4. Calidad de datos

La literatura especializada destaca que la reutilización de datos provenientes de registros electrónicos de salud requiere una evaluación sistemática de su calidad. La calidad de los datos es esencial para garantizar la validez de los análisis basados en registros clínicos, especialmente en estudios predictivos. Según Weiskopf y Weng (2013), la calidad debe evaluarse considerando cinco dimensiones clave: completitud, corrección, concordancia, plausibilidad y vigencia (43).

Dado que errores en la integridad, consistencia o completitud pueden sesgar estimaciones, aumentar la incertidumbre y deteriorar el rendimiento real de los modelos, se implementó un proceso sistemático de evaluación de calidad de datos, siguiendo principios establecidos en la literatura de ingeniería de datos, informática médica y ciencias de la salud.

Con este propósito, se desarrolló una estrategia estructurada basada en el paradigma de validación por expectativas, enfoque ampliamente utilizado en herramientas modernas de aseguramiento de calidad como *Great Expectations* (GX), que es un *framework* de validación de datos ampliamente utilizado para garantizar la calidad, consistencia y

confiabilidad de los *dataset* antes de su uso analítico. Su enfoque se basa en la definición explícita de “expectativas”, es decir, reglas formales que describen el comportamiento esperado de cada variable, como rangos válidos, formatos, dominios o relaciones entre columnas.(44)

Siguiendo este marco conceptual, las validaciones se implementaron mediante procedimientos programáticos en Python, estructurados según los principios de la metodología GX, la cual permite definir expectativas explícitas y reproducibles sobre el comportamiento esperado de las variables. Las evaluaciones se organizaron en tres dimensiones analíticas principales: completitud, plausibilidad de valores (rangos) y consistencia de dominios categóricos.

- Evaluación de completitud: La completitud se refiere a la proporción de datos efectivamente registrados en relación con el total esperado para cada variable. En investigaciones clínicas, este aspecto es fundamental, ya que los valores faltantes pueden ser indicativos de inconsistencias en el registro, variaciones entre unidades de atención, fallas en los sistemas de captura o problemas de integración entre las distintas fuentes de datos. Evaluar la completitud permite identificar variables críticas con niveles inadmisibles de ausencia, orientar estrategias de imputación y prevenir sesgos derivados del análisis de datos incompletos.
- Evaluación de rangos plausibles: La evaluación de rangos plausibles consiste en verificar que los valores numéricos se ubiquen dentro de intervalos fisiológicos, administrativos o clínicamente razonables. Este control es esencial para detectar errores de digitación, cifras incompatibles con la fisiología humana, valores productos de fallas del equipamiento o registros imposibles desde el punto de vista asistencial. La validación de rangos asegura que el *dataset* mantenga coherencia interna y evita que datos aberrantes distorsionen los análisis descriptivos y afecten el desempeño de los modelos predictivos. Este enfoque es especialmente relevante en variables como signos vitales, edad, parámetros clínicos y valores derivados de mediciones instrumentales.
- Evaluación de dominios y consistencia categórica: La evaluación de dominios y consistencia categórica tiene por objetivo garantizar que las variables clasificadas como



categorías presenten únicamente valores válidos dentro de un conjunto predefinido. Este análisis permite detectar categorías erróneas, etiquetas inconsistentes, codificaciones obsoletas o variaciones en la nomenclatura que puedan comprometer la interpretación de la información. La consistencia categórica es crucial en registros clínicos donde categorías como *triage*, sexo, previsión, códigos diagnósticos o unidades de ingreso deben mantenerse uniformes y estandarizadas. Validar los dominios asegura que las categorías sean comparables y semánticamente coherentes, evitando problemas en la generación de tablas, cruces estadísticos y modelos predictivos basados en variables nominales.

### 6.5. Definición operacional de *outlier* superior:

Los *outliers* superiores corresponden a aquellos pacientes cuya estancia hospitalaria excede el punto de corte superior definido para su respectivo GRD.

Este umbral se calcula mediante la fórmula:

$$Outliers_{superiores} = (P_{75}) + (1,5 \times [P_{75} - P_{25}])$$

Donde P75 y P25 representan los percentiles 75 y 25 de la distribución de estancias para ese GRD.

Este criterio se encuentra estandarizado en la Norma Técnica MINSAL 2018–2019 para GRD-IR(9), que establece dicha fórmula estadística para *outliers* superiores.

### 6.6. Preprocesamiento de datos

El preprocesamiento de datos se realizó mediante un protocolo sistemático para asegurar la calidad y coherencia de la información. Se identificaron y corrigieron errores, duplicados y valores imposibles, estandarizando variables demográficas, administrativas y clínicas. Se verificaron rangos fisiológicos, se homogenizaron categorías y se manejaron datos faltantes mediante imputación o exclusión. Con *pandas* y *numpy* se normalizaron y codificaron las variables, generando además nuevas categorías derivadas para mejorar la interpretabilidad. Posteriormente, se aplicó un análisis descriptivo y comparativo entre *inliers* y *outliers*, complementado con cálculos de *odds ratios* e intervalos de confianza, identificando factores asociados al riesgo de estancia prolongada.

## 6.7. Modelos predictivos

Con el propósito de estimar la probabilidad de que un paciente se constituya en un *outlier* superior desde el momento de su ingreso hospitalario, se implementaron modelos supervisados de aprendizaje automático (machine learning). La predicción temprana de *outliers* superiores representa un desafío metodológico relevante, dada la naturaleza multifactorial de los procesos asistenciales y el marcado desbalance en la distribución de clases. Se describen los algoritmos utilizados, la estrategia de entrenamiento y validación temporal, los criterios de selección de variables y las métricas empleadas para evaluar el rendimiento de cada modelo.

- Algoritmos de Aprendizaje Automático Evaluados: Se evaluó un conjunto de algoritmos supervisados, seleccionados por su uso extendido en predicción clínica y por su robustez ante diversos tipos de datos. Los modelos implementados fueron:
  - Regresión Logística
  - Random Forest
  - XGBoost
  - LightGBM
  - Support Vector Classifier (SVC)
  - Naive Bayes
  - Red Neuronal Multicapa (MLP)

El problema presenta un desequilibrio notable de clases, con una proporción sustancialmente menor de pacientes clasificados como *outliers* superiores. Para enfrentar esta situación y mejorar la sensibilidad hacia la clase minoritaria, todos los algoritmos fueron implementados en sus versiones ajustadas para datos desbalanceados (*imbalance-aware*). Según el algoritmo, se utilizaron estrategias como *class\_weight*, *scale\_pos\_weight*, *sample\_weight* o *priors*, lo que permitió penalizar de manera diferencial los errores en la clase minoritaria y evitar el sesgo hacia la clase mayoritaria.

- Estrategia de Entrenamiento y Validación Temporal: El proceso de entrenamiento utilizó los egresos hospitalarios correspondientes a los años 2018 y 2019, los cuales representan el 63,3 % del total de registros disponibles y corresponden a un período pre-pandemia

caracterizado por una distribución estable de diagnósticos, flujos asistenciales y condiciones operativas del hospital.

La cohorte de validación se conformó con los datos del año 2022, que representan el 36,7 % de los registros. Este período posterior a la pandemia de COVID-19 se caracteriza por cambios epidemiológicos, organizacionales y en la demanda asistencial, lo que constituye un conjunto temporalmente independiente y exigente para evaluar la capacidad de generalización del modelo.

La elección de esta estrategia de validación temporal responde a la necesidad de simular un escenario real de implementación, donde el modelo se entrena con datos históricos y debe mantener su desempeño ante pacientes de períodos futuros. Esta separación estricta por año evita el sobreajuste y previene el data *leakage*, permitiendo evaluar la estabilidad del modelo frente a variaciones en la casuística y en la complejidad de los episodios hospitalarios.

- Selección de Variables: La selección de predictores se realizó mediante un método de *Forward Selection*, que evalúa secuencialmente el aporte incremental de cada variable al rendimiento del modelo. Este enfoque permite construir modelos más parsimoniosos, reduciendo la complejidad, la redundancia y la colinealidad entre predictores, mientras se preserva la capacidad explicativa y predictiva del conjunto final de variables. Este método permitió identificar cuáles características aportaron mayor valor predictivo al modelo y cuáles generaban un efecto marginal o redundante.
- Métricas de Desempeño: Para evaluar el rendimiento de los modelos se utilizaron las métricas más relevantes en contextos con desbalance de clases, entre ellas:
  - Precisión (*Precision*): proporción de predicciones positivas correctas.
  - Sensibilidad (*Recall*): capacidad del modelo para identificar correctamente a los *outliers* superiores (clase minoritaria).
  - F1-score: medida armónica entre precisión y sensibilidad, útil cuando existe asimetría en la distribución de clases.
  - Área Bajo la Curva ROC (AUC-ROC): indicador global del rendimiento discriminativo del modelo.

Estas métricas permiten un análisis integral tanto del comportamiento global como del desempeño específico sobre la clase de mayor interés clínica.

- Tratamiento del Desbalance de Clases: Debido a que los *outliers* superiores representan únicamente el 4,6 % de la cohorte total, se implementaron técnicas de sobre-muestreo para mitigar el sesgo inherente hacia la clase mayoritaria. Se utilizaron métodos ampliamente validados en la literatura, tales como:
  - SMOTE (*Synthetic Minority Over-sampling Technique*)
  - SMOTEENN (*SMOTE + Edited Nearest Neighbors*)
  - ADASYN (*Adaptive Synthetic Sampling*)

Estas técnicas generan muestras sintéticas o combinadas de la clase minoritaria, reduciendo la tendencia de los modelos a predecir exclusivamente la clase mayoritaria y mejorando la capacidad de detección de los pacientes con riesgo aumentado de estancia prolongada.

El análisis de explicabilidad se apoyó en la interpretación de los coeficientes, probabilidades condicionales o pesos internos según el algoritmo, permitiendo describir de qué manera las variables seleccionadas influyen en la probabilidad de que un paciente se clasifique como *inlier* u *outlier* superior. Este enfoque orientado a la importancia de los predictores proporciona una comprensión transparente del comportamiento del modelo y facilita la interpretación clínica de los factores asociados a estancias hospitalarias prolongadas.

#### **6.8. Aspectos éticos y manejo de datos:**

El estudio se desarrolló en estricto cumplimiento de las políticas institucionales del HCUCH relativas a la protección de datos personales, confidencialidad de la información clínica y resguardo ético de los participantes. Para su ejecución, se utilizaron exclusivamente datos previamente anonimizados, sin posibilidad de identificación directa o indirecta de los pacientes. Asimismo, el protocolo fue revisado y aprobado por el Comité de Ética Científico del HCUCH, contando con su autorización formal emitida el 14 de junio de 2023, fecha en que se otorgó la aprobación ética correspondiente para la realización de la investigación (Anexo N° 03).

## 6.9. Software y entorno

El análisis cuantitativo y el procesamiento de los datos se realizaron empleando Python 3.11, junto con un conjunto de bibliotecas especializadas ampliamente utilizadas en investigación científica y en el desarrollo de modelos de aprendizaje automático. Para la gestión y manipulación de los datos se utilizaron *pandas* y *numpy*, mientras que *matplotlib* y *seaborn* se emplearon para la generación de visualizaciones descriptivas y análisis exploratorios. La construcción, entrenamiento y evaluación de modelos clásicos de machine learning se efectuó mediante *scikit-learn*, y el desarrollo de arquitecturas basadas en redes neuronales se llevó a cabo utilizando *TensorFlow*.

Todo el flujo de trabajo analítico, desde la depuración de los datos hasta la implementación de los modelos predictivos, se ejecutó en el entorno de desarrollo Visual Studio Code (VS Code) versión 2, el cual facilitó una integración eficiente de las herramientas de programación, control de versiones y visualización, asegurando un proceso reproducible, ordenado y metodológicamente coherente.

## **7. RESULTADOS**

### **7.1. Comprender el proceso de clasificación GRD del HCUCH (OE1)**

El proceso de asignación del código GRD a un paciente hospitalizado constituye una etapa clave para la estandarización y evaluación de los episodios de atención. Este procedimiento se inicia desde el momento mismo de la admisión del paciente al hospital, instancia en la que se registran los datos administrativos esenciales, tales como edad, sexo, fecha de ingreso y servicio clínico responsable. Paralelamente, durante la hospitalización se genera y consolida información clínica mediante sistemas institucionales, destacando el RCE en el HCUCH, el sistema TiCares y la hoja de alta médica (epicrisis), donde se documentan diagnósticos, procedimientos, evolución clínica y condiciones al egreso. Una vez completado el proceso de hospitalización, los registros administrativos y clínicos son integrados, permitiendo la generación final del código GRD correspondiente a cada episodio (Figura N°03).

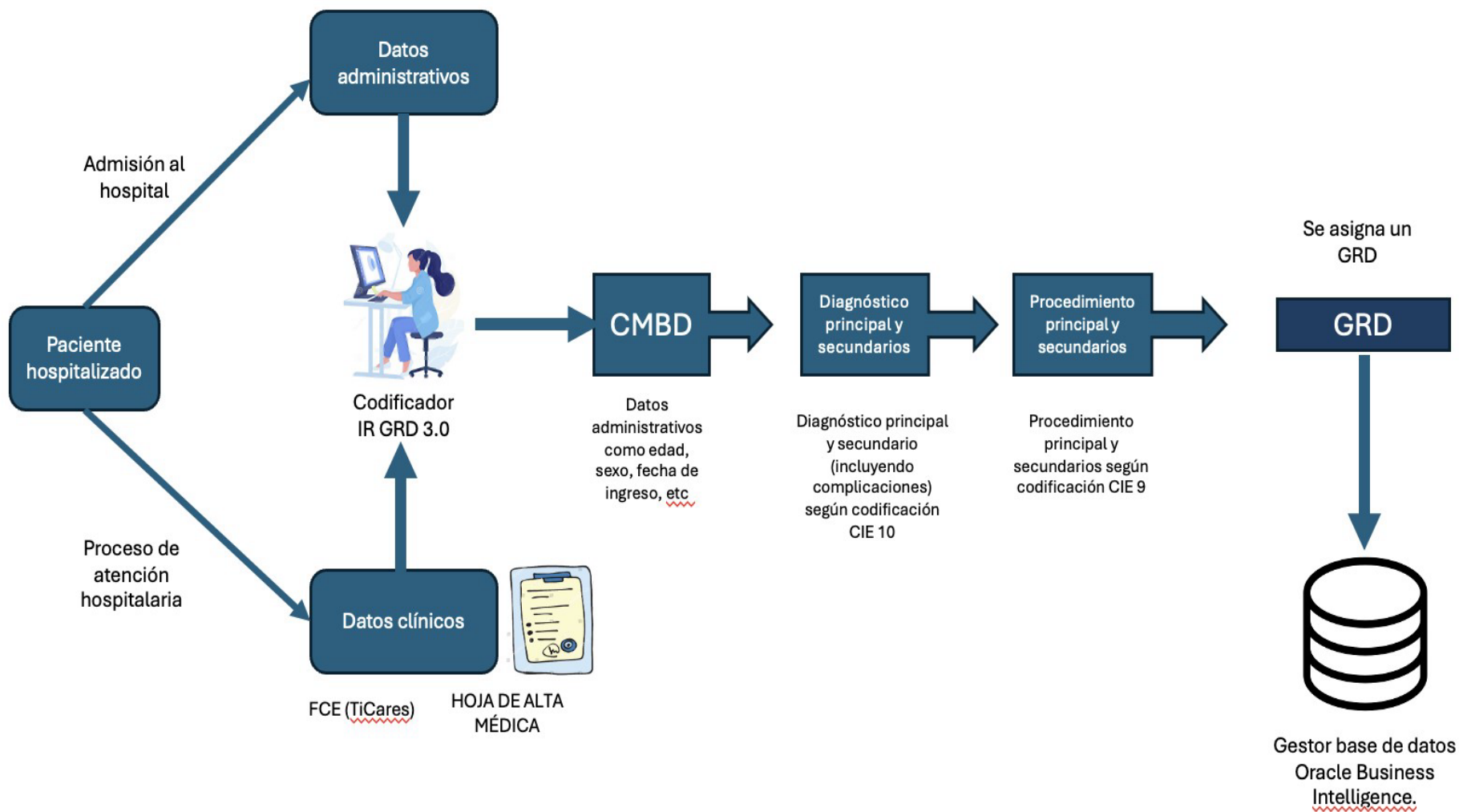
Los algoritmos de agrupación GRD tienen su origen en un marco conceptual común, evolucionando progresivamente desde los HCFA-DRGs desarrollados en 1983 en Estados Unidos, hasta los sistemas IR-DRGs (International Refined Diagnosis Related Groups) introducidos en el año 2000. A lo largo del tiempo, estos sistemas han sido adaptados para responder a las necesidades epidemiológicas, organizacionales y financieras de distintos países. En Chile, el sistema adoptado es el IR-GRD versión 3.0, utilizado tanto en el sector público como privado, y que permite codificar episodios de atención ambulatoria y hospitalaria según una clasificación clínicamente coherente y comparable a nivel internacional.

La correcta asignación de un GRD requiere la integración de datos administrativos y datos clínicos, los cuales alimentan el Conjunto Mínimo Básico de Datos (CMBD). Este conjunto constituye la estructura central donde se recopila información clave del episodio, incluyendo diagnósticos principales y secundarios registrados según la Clasificación Internacional de Enfermedades, CIE-10, así como la identificación de complicaciones ocurridas durante la hospitalización. Del mismo modo, se codifican los procedimientos médicos efectuados, utilizando la clasificación CIE-9-MC, lo que permite caracterizar la complejidad asistencial y la intensidad del uso de recursos.

Una vez consolidada toda esta información, el sistema de agrupación determina el GRD final, el cual sintetiza el episodio hospitalario desde una perspectiva clínica y administrativa, atribuyendo a cada caso una categoría que refleja su diagnóstico principal, comorbilidades, severidad y procedimientos asociados. Esta clasificación no solo permite comparar episodios entre pacientes, unidades o instituciones, sino que constituye una herramienta fundamental para la gestión hospitalaria, la planificación de recursos y la evaluación de desempeño.

Finalmente, el GRD asignado es almacenado en plataformas de gestión de datos, como *Oracle Business Intelligence*, donde queda disponible para análisis clínico, auditorías, estudio de indicadores, proyecciones financieras y procesos de mejora continua. Este flujo de información estructurado garantiza la estandarización de la codificación, fortalece la calidad del registro clínico y contribuye a una toma de decisiones informada en los distintos niveles de gestión en salud.

Fig N° 03: Proceso de obtención de la codificación GRD.





## **7.2. Describir la base de datos GRD y caracterizar sus variables, identificando patrones asociados a estancia hospitalaria prolongada (OE2).**

### **7.2.1. Descripción base de datos GRD y DAU del HCUCH**

La base de datos GRD y DAU analizado, corresponde a una base de datos estructurada de tipo tabular. Está conformado por 13.760 registros (observaciones) y 153 atributos (variables), lo que refleja un volumen considerable de información clínica y administrativa asociada a episodios hospitalarios.

El examen de su estructura interna, obtenido mediante la función `df.info()`, evidencia la presencia de 27 variables de tipo entero (`int64`), 35 variables numéricas de punto flotante (`float64`) y 91 variables de tipo objeto (`object`). Estas últimas corresponden principalmente a datos textuales, categorías codificadas y campos derivados de registros clínicos o administrativos, incluyendo posibles formatos de fecha.

### **7.2.2. Evaluación de la calidad de datos**

Para la evaluación de la calidad de los datos se empleó la metodología proporcionada por *Great Expectations* (GX), una librería de Python orientada a la validación automatizada y estandarizada de conjuntos de datos. Este enfoque permitió definir y aplicar reglas explícitas de calidad sobre las variables incluidas en el estudio, organizadas en estas dimensiones: completitud (presencia de datos esperados), plausibilidad de rangos (verificación de valores dentro de intervalos fisiológicos o administrativos coherentes), evaluación de dominios (corroboración de que las categorías corresponden a conjuntos válidos y predefinidos) y consistencia categórica (uniformidad y coherencia interna en etiquetas y clasificaciones). La aplicación sistemática de estas validaciones facilitó la identificación de anomalías, la estandarización del *dataset* y la garantía de su integridad previa al análisis estadístico y predictivo.

- Evaluación de completitud: Se evaluaron variables demográficas (sexo, edad, nacionalidad y comuna), administrativas (servicio de ingreso y tipo de previsión) y clínicas (frecuencia cardíaca, saturación de oxígeno, presión arterial, frecuencia respiratoria, puntaje de Glasgow). Para cada variable se calculó:
  - Número absoluto de datos faltantes.

- Porcentaje relativo respecto del total
- Cumplimiento de un umbral de aceptabilidad de  $\leq 5\%$ , definido siguiendo recomendaciones de la literatura en ciencia de datos clínicos y análisis de registros electrónicos.

La validación se implementó mediante reglas automatizadas equivalentes a la expectativa *expect\_column\_values\_to\_not\_be\_null* de GX, permitiendo detectar variables con potencial riesgo de sesgo o que requerían imputación, revisión de origen o exclusión.

- Evaluación de rangos plausibles: Para garantizar que los valores registrados fueran consistentes con la fisiología humana y con las características esperadas del proceso asistencial, se establecieron intervalos de plausibilidad para variables numéricas. Estos rangos se definieron en base a literatura clínica, parámetros fisiológicos normalizados y experiencias metodológicas de auditoría de datos en hospitales de alta complejidad.

Ejemplos de rangos definidos:

- Edad: 15–110 años
- Frecuencia cardíaca: 20–220 lpm
- Frecuencia respiratoria: 4–60 rpm
- Saturación de oxígeno: 50–100%
- Presión arterial sistólica: 50–260 mmHg
- Glasgow: 3–15 puntos

Los valores fuera de estos intervalos fueron clasificados como inválidos y cuantificados. La regla aplicada fue equivalente al método *expect\_column\_values\_to\_be\_between* utilizado por GX. Este procedimiento permitió identificar errores de digitación, artefactos de monitorización, registros incompletos y valores no plausibles que requerían revisión o depuración.

- Evaluación de dominios y consistencia categórica: Se verificó que las variables categóricas contuvieran exclusivamente valores pertenecientes a conjuntos predefinidos, asegurando integridad semántica y coherencia analítica. Estas variables

son fundamentales para la estratificación y el modelamiento predictivo, por lo que su validación es esencial. Los dominios evaluados incluyeron:

- outlier\_sup: {0, 1}
- triage: {I, II, III, IV, V}
- año: {2018, 2019, 2022}
- sexo: {1, 2} (equivalentes a masculino/femenino según codificación institucional)

La regla empleada fue análoga a *expect\_column\_values\_to\_be\_in\_set* de GX. Este paso permitió identificar inconsistencias de codificación, errores en la integración de fuentes o categorías no reconocidas que podían afectar el modelado posterior.

- Integración y reporte sistematizado: Tras ejecutar las validaciones en las tres dimensiones anteriores, se consolidaron los resultados en un reporte unificado que clasificó cada variable según:
  - cumplimiento o incumplimiento de las reglas definidas,
  - porcentaje de completitud,
  - número y proporción de valores fuera de rango,
  - porcentaje de categorías inválidas.

La Tabla N° 02 muestra un resumen estructurado de la calidad de los datos, organizado en tres dimensiones fundamentales: completitud, rangos plausibles y dominio/consistencia categórica, siguiendo una metodología de validación similar a la implementada mediante GX. Se registró una alta completitud en los datos, lo cual es fundamental para evitar sesgos, pérdida de información y necesidad excesiva de imputación. La presencia de valores fuera de rango sugiere errores de digitación, fallas de equipos, registros mal ingresados o situaciones clínicas extremas, lo que requiere corrección, exclusión o verificación antes del análisis. Las variables categóricas presentan excelente integridad semántica, sin codificaciones erróneas ni categorías fuera de los valores esperados.

Tabla N° 02: Resumen de la evaluación de la calidad de datos

Variable de interés	Detalle	n° de datos faltantes	% de datos faltantes	n° de datos inválidos	% de datos inválidos
<b>SEGÚN COMPLETITUD</b>					
outlier_sup	NA ≤ 5.0%	0	0	0	0
sexo	NA ≤ 5.0%	0	0	0	0
edad	NA ≤ 5.0%	0	0	0	0
nacionalidad	NA ≤ 5.0%	4	0,03	0	0
comuna	NA ≤ 5.0%	2	0,01	0	0
servicio	NA ≤ 5.0%	0	0	0	0
tipo_prevision_ok	NA ≤ 5.0%	0	0	0	0
triage	NA ≤ 5.0%	0	0	0	0
pas	NA ≤ 5.0%	0	0	0	0
pad	NA ≤ 5.0%	0	0	0	0
pam	NA ≤ 5.0%	0	0	0	0
sat	NA ≤ 5.0%	0	0	0	0
fc	NA ≤ 5.0%	0	0	0	0
fr	NA ≤ 5.0%	0	0	0	0
glasgow	NA ≤ 5.0%	0	0	0	0
cod_urg_cie10	NA ≤ 5.0%	0	0	0	0
<b>SEGÚN RANGOS PLAUSIBLES</b>					
edad	[15,110]	0	0	0	0
fc	[20,220]	0	0	329	2,39
fr	[4,60]	0	0	1819	13,22
glasgow	[3,15]	0	0	930	6,76
pad	[10,200]	0	0	398	2,89
pam	[20,200]	0	0	377	2,74
pas	[50,260]	0	0	381	2,77
sat	[50,100]	0	0	352	2,56
<b>SEGÚN DOMINIO Y CONSISTENCIA CATEGÓRICA</b>					
ano	{2018,2019,2022}	0	0	0	0
outlier_sup	{0,1}	0	0	0	0
sexo	{1,2}	0	0	0	0
triage	{'I','II','III','IV','V'}	0	0	0	0

### 7.2.3. Preprocesamiento y análisis de datos

El preprocesamiento de los datos se desarrolló siguiendo un protocolo sistemático orientado a garantizar la calidad, coherencia y utilidad analítica de las variables incluidas en el estudio. En primer lugar, se efectuó una revisión inicial para identificar errores estructurales, valores imposibles, duplicados y discrepancias en los registros. Posteriormente, se estandarizaron los formatos de todas las variables demográficas (edad, sexo, nacionalidad, residencia), administrativas (servicio de ingreso, previsión) y clínicas (*triage*, presión arterial, frecuencia cardíaca, frecuencia respiratoria, saturación de oxígeno, Glasgow y diagnóstico CIE-10). Para las variables numéricas se verificaron rangos fisiológicos plausibles, corrigiendo o excluyendo valores inconsistentes según criterios clínicos. En las variables categóricas se homogeneizaron etiquetas y se agruparon categorías poco frecuentes para mejorar la estabilidad estadística.

El proceso de limpieza y normalización de datos se realizó utilizando las bibliotecas *pandas* y *numpy* de Python, las cuales permitieron gestionar eficientemente grandes volúmenes de información clínica y administrativa. El manejo de datos faltantes se realizó mediante el análisis de su patrón de ausencia, aplicando imputación o exclusión según el porcentaje y relevancia clínica de la variable. Asimismo, se generaron nuevas variables derivadas, por ejemplo, *triage* dicotomizado, previsión recodificada y región de residencia agrupada, con el fin de optimizar la interpretabilidad y reducir la colinealidad. Finalmente, se llevó a cabo una normalización y codificación adecuada para los algoritmos de aprendizaje automático, garantizando que todas las variables ingresaran al modelamiento en un formato estructurado, estandarizado y listo para análisis predictivo. Si quieres, puedo adaptarlo exactamente a tu base de datos real línea por línea.

Posteriormente, se llevó a cabo un análisis descriptivo univariado, orientado a caracterizar la distribución de las variables y comparar sistemáticamente a los pacientes clasificados como *inliers* versus aquellos identificados como *outliers* superiores. Para ello, se utilizaron medidas de tendencia central y dispersión, así como pruebas de hipótesis apropiadas según la naturaleza de las variables (pruebas t, U de Mann–Whitney, chi cuadrado o exacta de Fisher), con el fin de identificar diferencias estadísticamente significativas entre ambos grupos.

Adicionalmente, se exploró la fuerza de asociación entre variables clínicas, demográficas y administrativas mediante el cálculo de *odds ratios* (OR) y sus respectivos intervalos de confianza al 95%, complementado con la estimación de valores p para determinar la significancia estadística. Este enfoque permitió identificar factores relacionados con la probabilidad de constituirse como *outlier* superior, aportando evidencia cuantitativa relevante para la construcción posterior de los modelos predictivos.

#### **7.2.4. Transformaciones de variables**

Se seleccionaron 15 variables, que se encontraban disponibles al momento del ingreso hospitalario. Se clasificaron en 3 dominios: demográficos, administrativos y clínicos. Algunas variables fueron transformadas para mejorar su interpretación y comparabilidad (Tabla N° 03):

- En el dominio demográfico, la variable nacionalidad fue recodificada en dos categorías: chilena y extranjera. Asimismo, la variable comuna de residencia se transformó en una clasificación dicotómica que distingue entre Región Metropolitana y otras regiones, con el fin de simplificar su análisis y mejorar la interpretabilidad de los resultados.
- En el dominio administrativo, la variable servicio de ingreso fue recodificada agrupando las distintas unidades hospitalarias en categorías analíticamente más manejables. Por su parte, la variable tipo de previsión se transformó en una clasificación dicotómica compuesta por FONASA y Otras previsiones, con el propósito de simplificar su interpretación y facilitar su uso en los análisis posteriores.
- En el dominio clínico, se aplicaron transformaciones más específicas. La variable triage fue recodificada en dos grupos clínicamente relevantes: alto (niveles I y II, correspondientes a mayor gravedad) y no alto (niveles III, IV y V). Esta transformación binaria permite capturar el riesgo inicial del paciente de manera más simple y analíticamente manejable. Asimismo, la variable CIE-10 de urgencia fue reorganizada agrupando distintos códigos diagnósticos en categorías más amplias, según sistemas, síndromes o grupos etiológicos, con el fin de evitar una alta dispersión de categorías y mejorar la estabilidad estadística del modelo.

Tabla N° 03: Variables y transformaciones

Tipo de Variable	Variable de interés	Transformaciones
Demográficos	Sexo	No
	Edad	No
	Nacionalidad	Chileno/Extranjero
	Comuna	RM/Regiones
Administrativos	Servicio que ingresa desde la urgencia	No
	Tipo de previsión	FONASA/Otras
Clínicos	Triage	Alto (I y II) y no alto (III, IV, V)
	Presión arterial sistólica (PAS)	No
	Presión arterial diastólica (PAD)	No
	Presión arterial media (PAM)	No
	Frecuencia cardíaca (FC)	No
	Saturación de oxígeno (SatO2)	Sat baja (< 90%), Sat >= 90%
	Frecuencia respiratoria (FR)	No
	Escala de Glasgow	Glasgow bajo (<15), Glasgow 15
	Diagnóstico de urgencia CIE 10	No

### 7.2.5. Análisis descriptivo y univariado

Se realizó un análisis descriptivo de la base de datos GRD del HCUCH, considerando un total de 13.760 egresos hospitalarios codificados entre los años 2018, 2019 y 2022. En virtud de la heterogeneidad de los pacientes, enfermedades y capacidad de predecir la posibilidad de ser *outlier* superior con la información disponible al ingreso de hospitalización, se decidió que se incluirán los pacientes con edad igual o mayor de 15 años y que hayan ingresado por el servicio de urgencias del HCUCH.

Para identificar a los pacientes con estancias hospitalarias prolongadas, se utilizó como referencia el umbral superior estadístico definido:

$$\text{Estancia hospitalaria prolongada} = (P_{75}) + (1,5 \times [P_{75} - P_{25}])$$

Esta fórmula es una definición estática que corresponde al criterio establecido en la Norma MINSAL 2018–2019 GRD IR para la determinación de *outliers* superiores según cada GRD específico. La fórmula se aplicó a los egresos hospitalarios del HCUCH correspondientes a los años 2018, 2019 y 2022 de pacientes ingresados por el servicio de urgencia. Con base en este análisis que contribuye a sugerir factores de riesgo y protectores. Se consideró estancia prolongada aquella igual o superior a 23 días.

Aplicando este criterio a la base de datos, se identificó un subgrupo de pacientes con estancias prolongadas (estancia hospitalaria igual o mayor a 23 días), que incluyó 1224 casos, que representan el 8,9% de los casos.

En el conjunto de datos se identificaron 568 códigos GRD únicos, lo que refleja una marcada heterogeneidad diagnóstica en la población estudiada.

A continuación, se presentan las principales características de esta cohorte, organizadas en tres dimensiones: demográficas, administrativas y clínicas (Tabla N° 04)

Tabla N° 04: Comparación de variables demográficas, administrativas y clínicas según duración de la hospitalización (< 23 días vs ≥ 23 días).

Variables		EH < 23 días	EH ≥ 23 días	p-value
N		12.536	1.224	
Edad, median (IQR)		63 (46;75)	68 (57;78)	<0,001*
Sexo				0,711
	Masculino	6.326 (50,5%)	625 (51,1%)	
	Femenino	6.210 (49,5%)	599 (48,9%)	
Nacionalidad				0,033*
	Chileno	12.397 (98,9%)	1.202 (98,2%)	
	Extranjero	135 (1,1%)	22 (1,8%)	
Residencia				<0,001*
	R. Metropolitana	12.043 (96,1%)	1.135 (92,7%)	
	Otras regiones	493 (3,9%)	89 (7,3%)	
Tipo de previsión				<0,001*
	Fonasa	7.631 (60,9%)	815 (66,6%)	
	Isapre/Particular	4.905 (39,1%)	409 (33,4%)	
Admisión				<0,001*
	UPC	4.017 (32,0%)	740 (60,5%)	
	Otro servicio	8.519 (68,0%)	484 (39,5%)	
Triaje				<0,001*
	Alto (I y II)	5.363 (42,8%)	718 (58,7%)	
	No alto (III, IV, V)	7.173 (57,2%)	506 (41,3%)	
pas, median (IQR)		132 (115;150)	124 (104;147)	<0,001*
pad, median (IQR)		78 (67;89)	72 (60;85)	<0,001*
pam, median (IQR)		96.7 (84.3;108.3)	90 (75.7;105.3)	<0,001*
fc, median (IQR)		88 (74;103)	93 (77.8;110)	<0,001*
fr, median (IQR)		18 (16;18)	18 (16;20)	<0,001*
sat, median (IQR)		97 (95;98)	96 (93;98)	<0,001*
glasgow, median (IQR)		15 (15;15)	15 (15;15)	<0,001*

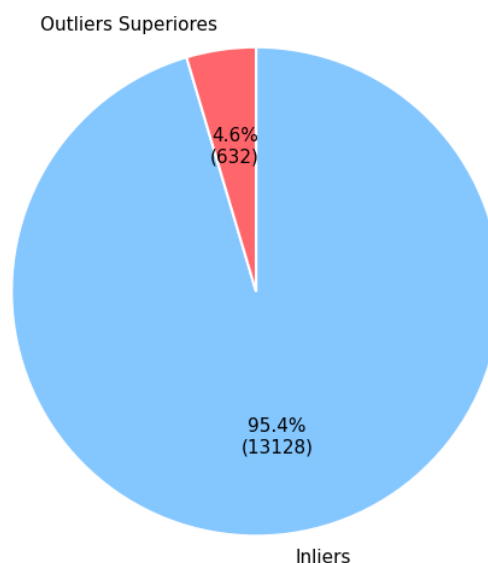


### 7.3. Identificar factores de riesgo asociados a *outliers* superiores mediante análisis conjunto de la base de datos GRD y RCE del HCUCH (OE3).

#### 7.3.1. Análisis de los *outliers* superiores e *inliers*

Del total de 13.760 casos analizados, 632 pacientes fueron identificados como *outliers superiores*, lo que corresponde al 4,6% (Figura N° 04). Estos representan los egresos cuya duración de hospitalización excedió significativamente los valores esperados según su grupo GRD. Por otro lado, 13.128 pacientes, lo que equivale al 95,4% fueron clasificados como *inliers*, es decir, su duración de estadía se mantuvo dentro del rango considerado normal. Esta distribución evidencia que los *outliers superiores* constituyen una minoría dentro de la población hospitalaria, lo que plantea desafíos metodológicos para su análisis y predicción, especialmente en el contexto de modelos de clasificación con clases desbalanceadas.

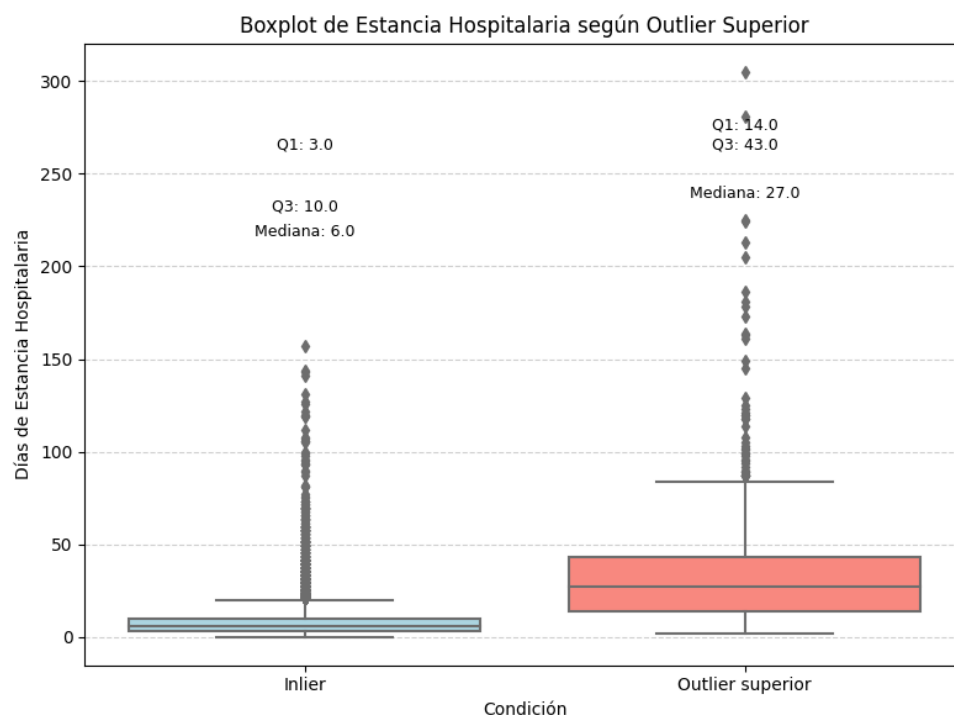
Fig N° 04: Distribución de *inliers* y *outliers superiores*. HCUCH. 2018-1019-2022



La duración de la estancia hospitalaria (en días) difiere entre pacientes clasificados como *outliers superiores* y pacientes *inliers*. Los *outliers superiores* tienen una mediana de estancia hospitalaria de 27 días, con un rango intercuartílico de 14 a 43 días, observándose además valores extremos considerablemente más altos en este grupo. En contraste, los pacientes *inliers* presentan una mediana de 6 días, con un rango intercuartílico de 3 a 10 días. Los resultados revelan una diferencia

estadísticamente significativa entre ambos grupos, con un valor  $p < 0,001$ . Esta diferencia sustancial en los días de hospitalización refuerza la definición operacional de *outliers superiores* como casos con estancia prolongada (Figura N° 05).

Fig N° 05: Mediana estancia hospitalaria de *inliers* y *outliers superiores*.  
HCUCH. 2018-1019-2022



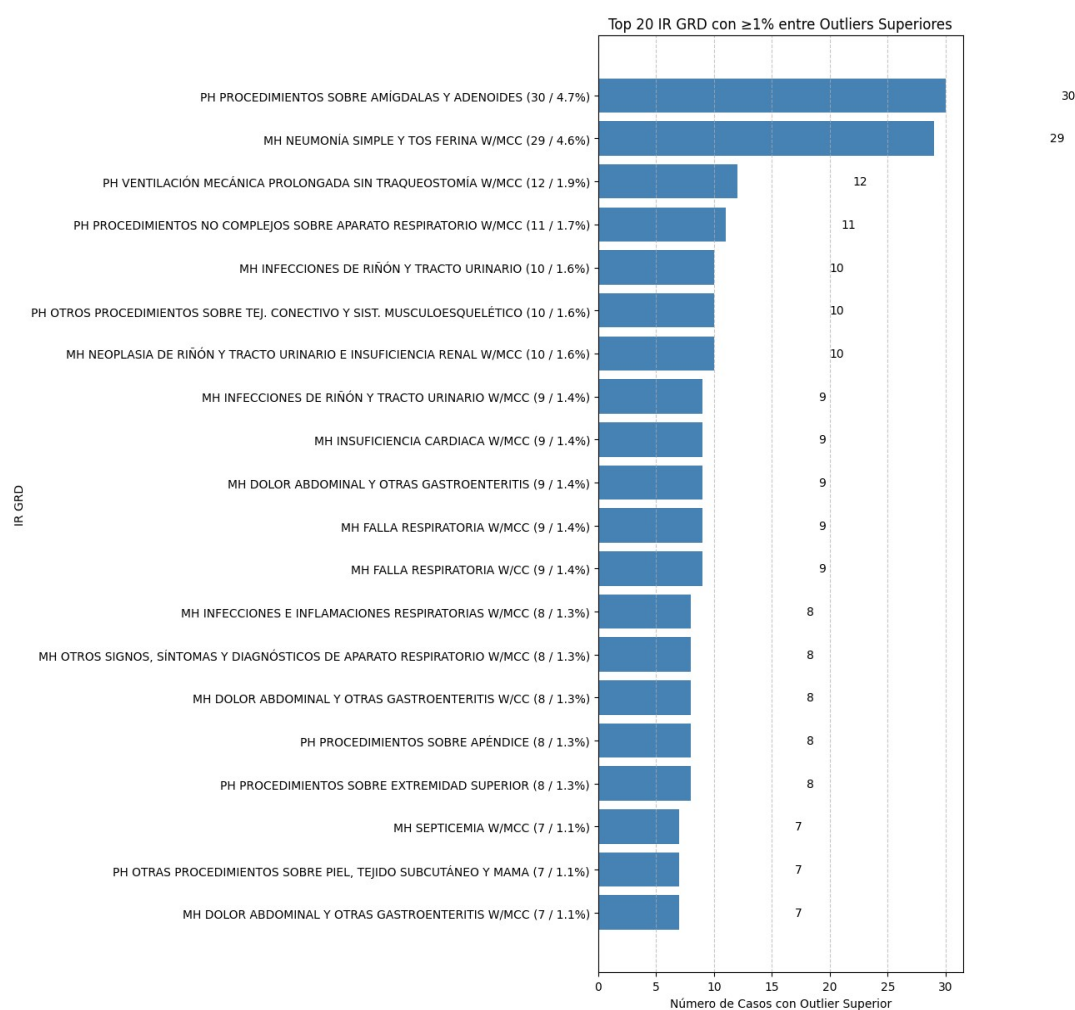
Con respecto al punto de corte superior de los GRD, es decir, el umbral en días de hospitalización utilizado para definir a los *outliers superiores*, se observa una distribución marcadamente asimétrica hacia la derecha. La mayoría de los puntos de corte se concentra en estancias menores a 40 días. A partir del día 50 la distribución cae de manera considerable, quedando solo casos aislados y dispersos. Pese a ello, se identifican valores atípicos de gran magnitud, con puntos de corte que alcanzan hasta 160 días.

### 7.3.2. Análisis de los GRD con mayor *outliers superiores*

La figura N° 06 se muestra los GRDs que concentran el mayor número de *outliers superiores*. Los tres GRDs con mayor frecuencia de *outliers superiores* son “PH procedimientos sobre amígdalas y adenoides” con 30 casos, lo que representa el 4,7

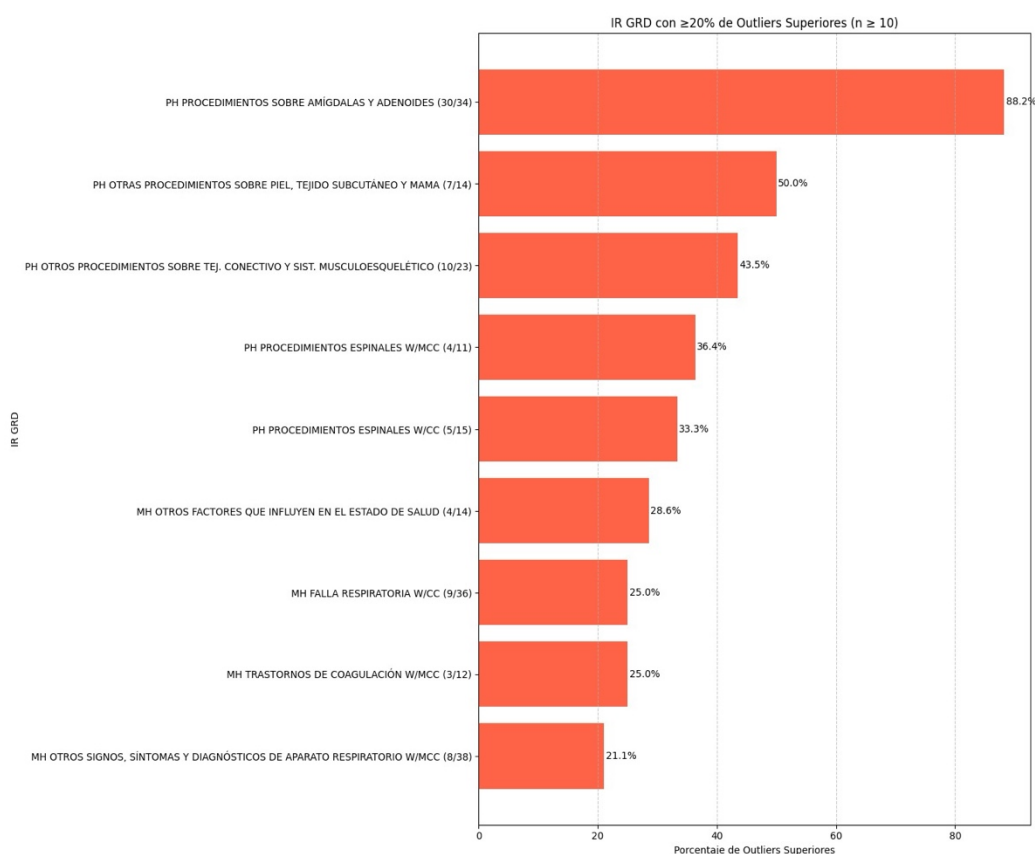
% de todos los *outliers superiores*, sigue “MH neumonía simple y tos ferina con comorbilidades mayores (W/MCC)” con 29 casos (4,6 %) y “PH ventilación mecánica prolongada sin traqueostomía W/MCC” con 12 casos (1,9 %). También destacan otros diagnósticos clínicos y quirúrgicos con entre 7 y 11 casos cada uno, como infecciones urinarias, insuficiencia cardíaca, dolor abdominal, falla respiratoria, infecciones respiratorias, septicemia, entre otros. Esta distribución evidencia que no son necesariamente los GRDs con mayor volumen total de casos los que concentran más *outliers superiores*, sino aquellos con mayor propensión a estancias hospitalarias prolongadas dentro de su categoría.

Fig N° 06: Distribución de los 20 IR-GRD con mayor número de casos clasificados como *outliers superiores* ( $\geq 1$  % del total), expresado en frecuencia absoluta y porcentaje.



La figura N° 07 muestra los GRDs que presentan un porcentaje igual o superior al 20 % de *outliers* superiores, considerando únicamente aquellos con al menos 10 casos totales. El GRD con mayor proporción de *outliers superiores* es “PH procedimientos sobre amígdalas y adenoides” en el que 30 de los 34 casos (88,2 %) superaron el punto de corte superior de su GRD, luego “PH otros procedimientos sobre piel, tejido subcutáneo y mama” con 50,0 % (7 de 14 casos) y “PH otros procedimientos sobre tejido conectivo y sistema musculoesquelético” con 43,5 % (10 de 23 casos). La relevancia de esto es que algunos GRD, aunque menos frecuentes, presentan una alta probabilidad de generar estancias prolongadas.

Fig N° 07: GRD con  $\geq 20$  % de casos clasificados como *outliers superiores* ( $n \geq 10$ ), según distribución porcentual.



La mayoría de los pacientes clasificados como *outliers superiores* tienen un exceso de días relativamente bajo; cerca de la mitad acumula hasta 5 días adicionales. Sin embargo, también se observan casos extremos con excesos muy prolongados, incluso

superiores a los 100 o 200 días, aunque son infrecuentes. A medida que aumenta el número de días como *outlier* superior, la frecuencia disminuye progresivamente, lo que indica una distribución fuertemente sesgada a la derecha.

### 7.3.3. Factores de riesgo asociados a *outliers* superiores

El análisis de las variables disponibles de forma precoz muestra las siguientes asociaciones con la condición de *outlier* superior (Tabla N° 05):

- Edad: Los pacientes clasificados como *outliers* superiores presentan una mediana de edad más alta (66 años [RIC 50;79]) en comparación con los *inliers* (63 años [RIC 47;76]), diferencia que resulta estadísticamente significativa ( $p = 0,002$ ).
- Sexo: La proporción de mujeres fue más alta entre los *outliers* superiores (53,3 %) que entre los *inliers* (49,3 %), sin alcanzar significación estadística ( $p = 0,053$ ).
- Nacionalidad: No se observaron diferencias significativas entre ambos grupos. La gran mayoría de los pacientes eran chilenos (casi 99 % en ambos), con un valor  $p = 0,912$ , lo que indica que la nacionalidad no se asocia a la condición de *outlier* superior.
- Residencia: El 5,9 % de los pacientes *outliers* superiores provienen de regiones distintas a la Metropolitana, en comparación con el 4,2 % de los *inliers*. Esta diferencia se acerca a la significación estadística ( $p = 0,048$ ), lo que sugiere que residir fuera de la Región Metropolitana podría estar relacionado con mayor riesgo de estancia prolongada.
- Tipo de previsión: No se observaron diferencias significativas ( $p = 0,193$ ). La proporción de pacientes FONASA fue algo mayor entre los *outliers* superiores (63,9 %) en comparación con los *inliers* (61,3 %).
- Triage: El 51,4 % de los *outliers* superiores fue clasificado con triaje de alta prioridad (niveles I–II), mientras que solo el 43,8 % de los *inliers* recibió esta clasificación. La diferencia fue estadísticamente significativa ( $p < 0,001$ ), lo que sugiere una asociación entre mayor severidad al ingreso y evolución hacia *outlier* superior.
- Presión arterial sistólica (PAS): No se observaron diferencias significativas en la mediana de la PAS entre ambos grupos (129 mmHg vs. 131 mmHg;  $p = 0,027$ ).

- Aunque estadísticamente significativa, la diferencia es pequeña y de relevancia clínica limitada.
- Presión arterial diastólica (PAD): Los *outliers* superiores presentaron valores significativamente menores de PAD (75 mmHg vs. 78 mmHg;  $p = 0,002$ ).
- Presión arterial media (PAM): La PAM fue también menor en los *outliers* superiores (94 mmHg vs. 96 mmHg;  $p = 0,006$ ).
- Frecuencia cardíaca (FC): La frecuencia cardíaca fue más alta en los pacientes *outliers* superiores (93 lpm [RIC 80;107]) en comparación con los *inliers* (88 lpm [RIC 74;103]), diferencia estadísticamente significativa ( $p < 0,001$ ).
- Frecuencia respiratoria (FR): Ambos grupos presentaron una mediana de 18 rpm, aunque la diferencia en la distribución fue significativa ( $p = 0,007$ ), lo que sugiere que los *outliers* tienen mayor tendencia a frecuencias respiratorias elevadas.
- Saturación de oxígeno (SatO<sub>2</sub>): Aunque la mediana fue similar en ambos grupos, los pacientes clasificados como *outliers* superiores presentaron saturaciones ligeramente menores (97 % [94–98]) en comparación con los *inliers* (97 % [95–98]). Esta diferencia, aunque pequeña, resultó estadísticamente significativa ( $p < 0,001$ ), lo que sugiere variaciones en la distribución de los valores entre ambos grupos.
- Escala de Glasgow: A pesar de que la mediana fue idéntica en ambos grupos (15 puntos), el análisis estadístico evidenció una diferencia significativa ( $p < 0,001$ ). Esto indica que las distribuciones completas de los puntajes difieren entre *outliers* superiores e *inliers*, aun cuando el valor central no cambia.

Tabla N° 05: Comparación de variables demográficas, administrativas y clínicas entre *inliers* y *outliers* superiores

Variables		<i>Inliers</i>	<i>Outlier superior</i>	p-value
N		13.128	632	
Edad, median (IQR)		63 (47;76)	66 (50;79)	0,002*
Sexo				
	Masculino	6.656 (50,7%)	295 (46,7%)	0,053
	Femenino	6.472 (49,3%)	337 (53,3%)	
Nacionalidad				0,912
	Chileno	12.975 (98,9%)	624 (98,7%)	
	Extranjero	149 (1,1%)	8 (1,3%)	
Residencia				0,048*
	R. Metropolitana	12.583 (95,8%)	505 (94,1%)	
	Otras regiones	545 (4,2%)	37 (5,9%)	
Tipo de previsión				0,193
	Fonasa	8.042 (61,3%)	404 (63,9%)	
	Isapre/Particular	5.086 (38,7%)	228 (36,1%)	
Triaje				<0,001*
	Alto (I y II)	5.756 (43,8%)	325 (51,4%)	
	No alto (III, IV, V)	7.372 (56,2%)	307 (48,6%)	
pas, median (IQR)		131 (115;150)	129 (110;149)	0,027*
pad, median (IQR)		78 (66;89)	75 (63;88)	0,002*
pam, median (IQR)		96.3 (83.7;108.0)	93.7 (80.6;107.3)	0,006*
fc, median (IQR)		88 (74;103)	92 (77;107)	<0,001*
fr, median (IQR)		18 (16;18)	18 (16;20)	0,007*
sat, median (IQR)		97 (95;98)	97 (94;98)	<0,001*
glasgow, median (IQR)		15 (15;15)	15 (15;15)	<0,001*

La Tabla N° 06 presenta la distribución de pacientes clasificados como *outliers* superiores según el servicio clínico de hospitalización al que fueron derivados desde urgencia. Los resultados evidencian que los servicios de Unidad de Pacientes Críticos (OR: 1,68), Otorrinolaringología (OR: 7,73) y Psiquiatría (OR: 4,75) mostraron *odds ratios* significativamente elevados, lo que indica una mayor probabilidad de que los pacientes ingresados a estas unidades se conviertan en *outliers* superiores. En contraste, los servicios de Cirugía (OR: 0,32), Unidad Coronaria (OR: 0,58) y Cardiología (OR: 0,34) presentaron *odds ratios* menores, reflejando una probabilidad reducida de constituirse como *outliers* superiores en comparación con el resto de los servicios clínicos.

Tabla N° 06: *Outliers* superiores según el servicio clínico de hospitalización al que ingresaron desde urgencia

Servicio	n	<i>Outlier</i> superior (%)	OR	p-valor
UNIDAD DE PACIENTES CRÍTICOS	4.757	293 (6,2%)	1,68	<0,001*
CIRUGIA	2.887	51 (1,8%)	0,32	<0,001*
MEDICINA INTERNA	1.949	89 (4,6%)	0,99	0,998
UNIDAD CORONARIA	831	23 (2,8%)	0,58	0,012*
TRAUMATOLOGIA	612	35 (5,7%)	1,28	0,207
NEFROLOGIA	529	31 (5,9%)	1,31	0,189
NEUROCIRUGIA	430	26 (6,0%)	1,35	0,178
GASTROENTEROLOGIA	390	13 (3,3%)	0,71	0,279
CARDIOLOGIA	371	6 (1,6%)	0,34	0,008*
NEUROLOGIA	322	15 (4,7%)	1,02	1
UROLOGIA	308	7 (2,3%)	0,48	0,067
GERIATRIA	171	6 (3,5%)	0,75	0,619
OTORRINOLARINGOLOGIA	122	32 (26,2%)	7,73	<0,001*
PSIQUIATRIA	27	5 (18,5%)	4,75	0,003*
PEDIATRIA	18	0 (0,0%)	0,58	0,713
HOSPITALIZACIÓN EN URGENCIA	15	0 (0,0%)	0,69	0,816
MEDICINA FISICA Y REHABILITAC.	13	0 (0,0%)	0,8	0,898

La Tabla N° 07 presenta los diagnósticos de ingreso por urgencia (según código CIE-10) con al menos 20 casos, identificando aquellos asociados a un mayor y menor riesgo de ser *outliers* superiores. Los diagnósticos con mayor probabilidad de convertirse en *outlier* superior fueron el absceso periamigdalino (OR: 16,23), seguido de artritis piógena (OR: 6,61), dolor (OR: 5,50), asma (OR: 4,39), historia personal de autolesión (OR: 4,18), insuficiencia respiratoria (OR: 2,14), gastroenteritis y colitis (OR: 2,57) y sepsis (OR: 1,97). En contraste, algunos diagnósticos se asociaron a un riesgo significativamente menor, entre ellos destacan apendicitis aguda (OR: 0,19), infarto agudo de miocardio (OR: 0,19), colecistitis (OR: 0,21), angina de pecho (OR: 0,20) y sangrado gastrointestinal (OR: 0,37), todos con *odds ratios* inferiores a 1.



Tabla N° 07: Diagnósticos de ingreso (CIE-10) asociados a mayor o menor riesgo de ser *outlier* superior

Diagnóstico (CIE10)	n	<i>Outlier</i> superior (%)	OR	p-valor
<b>MAYOR RIESGO</b>				
Sepsis (A41)	852	70 (8,2%)	1,97	<0,001*
Insuficiencia respiratoria (J96)	284	26 (9,2%)	2,14	<0,001*
Gastroenteritis y colitis (K52)	64	7 (10,9%)	2,57	0,033*
Absceso periamigdalino (J36)	63	27 (42,9%)	16,23	<0,001*
Historia personal de autolesión (Z91)	30	5 (16,7%)	4,18	0,006*
Artritis piógena (M00)	25	6 (24,0%)	6,61	<0,001*
Dolor (R52)	24	5 (20,8%)	5,5	<0,001*
Asma (J45)	23	4 (17,4%)	4,39	0,015*
<b>MENOR RIESGO</b>				
Apendicitis aguda (K35)	650	6 (0,9%)	0,19	<0,001*
Colecistitis (K81)	496	5 (1,0%)	0,21	<0,001*
Sangrado gastrointestinal (K92)	391	7 (1,8%)	0,37	0,010*
Angina de pecho (I20)	336	0 (0,0%)	0,03	<0,001*
Infarto agudo de miocardio (I21)	331	3 (0,9%)	0,19	0,002*

#### **7.4 Comparar modelos predictivos basados en algoritmos de aprendizaje automático que permita anticipar, desde la atención de urgencia, a los pacientes con mayor riesgo de transformarse en *outliers* superiores (OE4).**

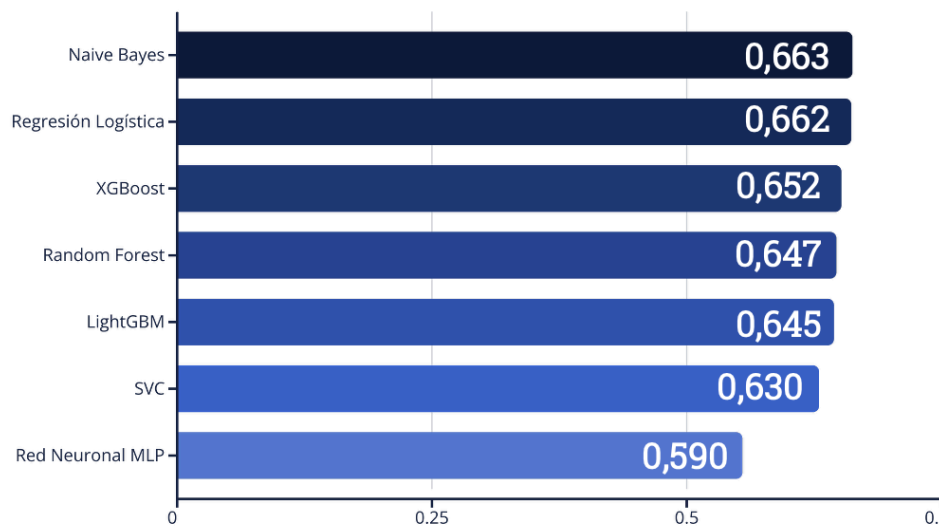
Para la elaboración de este trabajo se implementaron distintos algoritmos de clasificación, entre ellos Regresión Logística, Random Forest, XGBoost, LightGBM y Redes Neuronales Perceptrón Multicapa (MLP), utilizando las bibliotecas Scikit-learn y TensorFlow en Python, comparando su rendimiento mediante métricas de precisión, recall, F1-score y AUC ROC, en una cohorte de validación independiente (año 2022).

En todos los casos, el entrenamiento inicial se realizó con el total de 15 variables. Durante la fase inicial del modelado se identificó que varias de las variables presentaban un alto grado de colinealidad entre sí, lo que podía conducir a redundancia en la información aportada al modelo. Asimismo, se observó que ciertas variables exhibían una baja correlación con la variable dependiente, lo que sugería un aporte marginal en la capacidad explicativa. La inclusión de este tipo de predictores no solo podía introducir ruido estadístico, sino también incrementar de manera innecesaria los costos computacionales y, en consecuencia, afectar de forma negativa el desempeño y la capacidad de generalización del modelo. Con el objetivo de optimizar la parsimonia y mejorar la precisión predictiva, se implementó un proceso de selección de características mediante el método *Forward Selection*, el cual permitió priorizar aquellas variables con mayor relevancia y excluir las que no contribuían significativamente al modelo. El método *Forward Selection* corresponde a una técnica de selección secuencial de características utilizada en el ámbito del aprendizaje automático y la estadística. Su procedimiento consiste en iniciar con un modelo vacío, sin predictores, e ir incorporando progresivamente las variables que aportan mayor mejora al desempeño del modelo según un criterio predefinido. En cada iteración, se evalúa la inclusión de todas las variables candidatas restantes y se selecciona aquella que genere el mayor incremento en la capacidad predictiva. Este proceso se repite hasta que la adición de nuevas variables no aporte beneficios significativos al modelo. De esta manera, *Forward Selection* permite construir un conjunto parsimonioso de predictores, reduciendo la redundancia, mejorando la interpretabilidad y optimizando la eficiencia computacional.

Con el propósito de evaluar la capacidad de distintos algoritmos de aprendizaje supervisado para predecir la probabilidad de que un paciente se clasifique como *outlier* superior, es decir, que presente una estancia hospitalaria prolongada en relación con su Grupo Relacionado por Diagnóstico (GRD), se realizó un análisis comparativo del desempeño de siete modelos: Regresión Logística, Random Forest, XGBoost, LightGBM, Redes Neuronales (MLP), Support Vector Classifier (SVC) y Naive Bayes. Dado el marcado desbalance entre las clases (proporción considerablemente menor de pacientes *outliers* superiores respecto a los *inliers*), cada modelo se implementó utilizando su versión ajustada para datos desbalanceados (imbalance-aware), aplicando técnicas de ponderación interna o ajuste de pesos (`class_weight`, `scale_pos_weight`, `sample_weight` o `priors`) con el fin de optimizar la sensibilidad hacia la clase minoritaria y reducir el sesgo en la clasificación.

Los resultados obtenidos (Figura N° 08) muestran valores de AUC comprendidos entre 0.590 y 0.663, lo que evidencia una capacidad de discriminación moderada y relativamente homogénea entre los modelos. Los algoritmos Naive Bayes (AUC = 0.663) y Regresión Logística (AUC = 0.662) alcanzaron el mejor rendimiento global, seguidos de XGBoost (AUC = 0.652). En contraste, XGBoost, Random Forest y LightGBM presentaron sobreajuste hacia la clase mayoritaria, y las Redes Neuronales MLP el rendimiento más bajo (AUC = 0,590). Las métricas de desempeño más detalladas se pueden ver en la Tabla N° 08

Figura N° 08 Comparación de modelos predictivos de acuerdo a AUC ROC



Al haber un desbalanceo de datos, se aplicaron diversas técnicas de sobre-muestreo y combinación de clases, incluyendo SMOTE (Synthetic Minority Oversampling Technique), SMOTEENN y ADASYN (Adaptive Synthetic Sampling); sin embargo, estas no lograron mejorar el rendimiento global de los modelos.

Tabla N° 08: Comparación de métricas de desempeño (AUC, precisión, recall y F1score) entre los modelos predictivos aplicados.

Algoritmo	Variables	AUC		Precisión	Recall	F1-score
Regresión logística	7	0,653	Inlier	0,97	0,57	0,72
			Outlier	0,07	0,66	0,12
Random Forest	5	0,647	Inlier	0,98	0,30	0,47
			Outlier	0,05	0,87	0,10
XGBoost	5	0,652	Inlier	0,98	0,31	0,47
			Outlier	0,05	0,87	0,10
LightGBM	4	0,645	Inlier	0,98	0,30	0,46
			Outlier	0,05	0,88	0,10
RRNN MLP	1	0,555	Inlier	0,98	0,24	0,39
			Outlier	0,05	0,87	0,09
SVC	4	0,630	Inlier	0,96	1,00	0,98
			Outlier	0,00	0,00	0,00
Naive bayes	7	0,663	Inlier	0,97	0,86	0,91
			Outlier	0,09	0,32	0,14

Los modelos de Regresión Logística y Naive Bayes fueron entrenados utilizando un conjunto de siete variables clínicas y administrativas seleccionadas por su relevancia en la predicción temprana de estancias hospitalarias prolongadas. Estas variables incluyeron: servicio protector (*servicio\_protector*), diagnóstico de riesgo (*diag\_riesgo*), Glasgow bajo (*glasgow\_bajo*), diagnóstico protector (*diag\_protector*), triage alto (*triage\_alto*), servicio de riesgo (*servicio\_riesgo*) y saturación baja (*sat\_baja*). A partir de estas características, la figura N° 08 ilustra el desempeño de ambos modelos diferenciando entre la clase minoritaria (*outliers* superiores) y la clase mayoritaria (*inliers*).

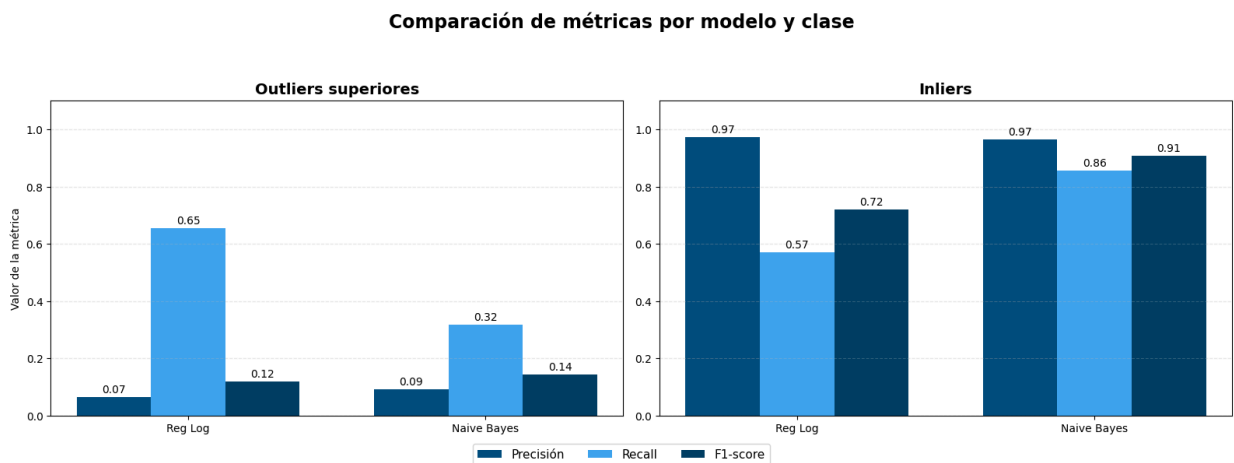
En la predicción de *outliers* superiores, los modelos muestran un rendimiento limitado, con valores de *precision* bajos y F1-score reducido. Esto sugiere que, aun cuando variables como *triage* alto o saturación baja aportan información temprana sobre la gravedad clínica, su capacidad aislada o combinada para anticipar estancias prolongadas es insuficiente para discriminar eficazmente esta clase minoritaria. No obstante, el *recall* observado en la Regresión Logística (0,65) indica que el modelo logra identificar una proporción relevante de casos verdaderamente prolongados.

En contraste, el desempeño para la clase *inlier* es considerablemente superior en ambos modelos. En el caso del modelo Naive Bayes, los resultados muestran que este algoritmo alcanzó un desempeño particularmente favorable en la predicción de *inliers*. Con un AUC de 0,663, el modelo exhibe una capacidad discriminativa comparable e incluso superior a la de otros clasificadores evaluados, pero destaca especialmente en las métricas asociadas a la clase mayoritaria. La *precision* para *inliers* (0,97) indica que casi todas las predicciones positivas para esta categoría fueron correctas, mientras que su *recall* (0,86) evidencia que el modelo logró identificar la gran mayoría de los pacientes que realmente pertenecían a esta clase. Como resultado, el F1-score de 0,91, el más alto entre todos los algoritmos analizados, confirma que Naive Bayes ofrece un equilibrio sobresaliente entre precisión y sensibilidad en la detección de *inliers*.

En suma, aunque las variables seleccionadas permiten un modelamiento robusto para la clase *inlier*, su capacidad para discriminar tempranamente a los *outliers* superiores resulta más limitada, lo cual es consistente con el carácter multifactorial y poco lineal

de las estancias prolongadas y con el marcado desbalance entre ambas clases en el conjunto de datos.

Fig N° 09: Comparación de métricas detalladas de los modelos Regresión Logística y Naive Bayes según *inliers* y *outliers* superiores



Desde una perspectiva de explicabilidad, el buen desempeño observado en el modelo Naive Bayes puede atribuirse a su estructura probabilística simple y altamente interpretable, la cual asume independencia condicional entre los predictores y estima la probabilidad de cada clase a partir de la contribución individual de cada variable. En el contexto de esta tesis, las variables seleccionadas, como triage alto, saturación baja, diagnóstico protector o servicio clínico, presentan patrones bien diferenciados para caracterizar a los pacientes con estancias dentro de lo esperado, lo que facilita que Naive Bayes identifique de forma eficiente combinaciones de atributos frecuentes en la clase *inlier*. Su mecanismo aditivo permite valorar explícitamente cómo cada predictor incrementa o reduce la probabilidad de pertenecer a dicha clase, evitando la sobreparametrización y minimizando el impacto de la colinealidad, elementos que afectan a modelos más complejos. Esta simplicidad estructural, sumada a su estabilidad frente al desbalance de clases, explica por qué Naive Bayes logra un rendimiento superior en la predicción de pacientes con estancias hospitalarias no prolongadas.

### 7.5 Identificar y analizar los factores que dificultan y limitan la capacidad predictiva de los modelos, comparados para la predicción de *outliers* superiores (OE5).

Después del análisis de los datos podemos nombrar los factores que dificultan la predicción de *outliers* superiores:

- La baja proporción de casos considerados *outliers* superiores, ya que sólo equivalía al 4,6 % de los casos, esto genera un marcado desbalance en la variable objetivo, provocando que los modelos tiendan a privilegiar la clase mayoritaria (*inliers*). Este fenómeno afecta de manera significativa la sensibilidad, dificultando la identificación de verdaderos positivos incluso en presencia de estrategias de ponderación o técnicas de sobremuestreo.
- Alta heterogeneidad de diagnósticos y procedimientos, ya que no se concentran en unos pocos GRD, sino que están distribuidos en múltiples categorías clínicas con diferentes perfiles (por ejemplo, infecciones respiratorias, procedimientos quirúrgicos, trastornos metabólicos, etc.). Esta fragmentación reduce la posibilidad de identificar patrones consistentes que puedan ser capturados por modelos supervisados, afectando la reproducibilidad y coherencia interna de las predicciones.
- Otro elemento crítico corresponde a la presencia de factores no observados al ingreso, los cuales ejercen un rol determinante en la prolongación de la estancia hospitalaria. Aunque se identificaron diferencias significativas en variables clínicas tempranas, como presión arterial, frecuencia respiratoria, nivel de conciencia o ingreso a UPC, muchos factores que condicionan estancias prolongadas (complicaciones intrahospitalarias, comorbilidades no registradas, calidad de cuidados, evolución clínica inesperada) no están disponibles en la etapa de urgencia, si no post admisión hospitalaria. La ausencia de estos determinantes limita intrínsecamente la capacidad del modelo para capturar el riesgo real de prolongación desde el inicio de la hospitalización.
- Solapamiento entre grupos en variables clínicas, aunque existen diferencias entre *inliers* y *outliers* en varios signos vitales (como PAM, FC, FR o Glasgow), los rangos intercuartílicos se superponen considerablemente, que indica la ausencia de umbrales fisiológicos que permitan una discriminación temprana clara,

dificultando la construcción de fronteras de decisión robustas por parte de los modelos supervisados.

- No existe relación no lineal entre GRD y días como *outlier* superior, que el exceso de días muestra una gran dispersión y asimetría en los días extra como *outlier*. La amplia dispersión, asimetría y variabilidad inter-GRD en los días excedentes sugiere que la prolongación de la estancia obedece a dinámicas complejas y multifactoriales, lo que desafía la capacidad de los modelos para capturar estas relaciones de manera estable.
- Finalmente, se identificaron GRD con altas proporciones de *outliers* pero con tamaños muestrales muy reducidos, lo que disminuye su capacidad de influir en modelos generalizados y puede introducir ruido en el proceso de aprendizaje cuando se incorporan de manera directa como predictores.

En contraste con estas limitaciones, los pacientes *inliers* presentan características que favorecen una predicción más estable y consistente. Al constituir la gran mayoría de los casos, su representación amplia permite que los modelos aprendan con mayor solidez los patrones asociados a estancias hospitalarias dentro del rango esperado. Además, los *inliers* exhiben una menor variabilidad clínica y una mayor homogeneidad diagnóstica, lo que facilita la identificación de relaciones más directas entre las variables registradas al ingreso y su evolución hospitalaria. La ausencia de complicaciones intrahospitalarias graves, la menor presencia de factores no observados y la mayor estabilidad en parámetros fisiológicos precoces reducen el solapamiento en las distribuciones y permiten que los modelos establezcan fronteras de decisión más claras. En conjunto, estas condiciones explican por qué, aun cuando los modelos fueron insuficientes para predecir de forma precisa a los *outliers* superiores, sí lograron identificar con alta certeza a los pacientes *inliers*.



## 8. DISCUSIÓN

El presente estudio permitió caracterizar integralmente el comportamiento de los pacientes con estancias hospitalarias prolongadas y *outliers* superiores en el HCUCH, así como evaluar la capacidad de modelos de aprendizaje automático para predecir *outliers* superiores al momento del ingreso hospitalario. Los resultados muestran que, pese a los avances logrados en la descripción del fenómeno y en la construcción de modelos iniciales, la predicción temprana de *outliers* superiores continúa siendo un desafío complejo que responde tanto a limitaciones estructurales del sistema GRD como a la naturaleza multifactorial del evento clínico estudiado.

En primer lugar, la caracterización del proceso de obtención del código GRD permitió situar el fenómeno dentro de un marco de clasificación retrospectiva que, si bien aporta estandarización y robustez administrativa, posee un rezago temporal inherente. Tal como se describe en la literatura clásica (45,46) los GRD fueron diseñados para fines de financiamiento y gestión, no para apoyar decisiones clínicas prospectivas, lo cual explica su limitada utilidad para anticipar estancias prolongadas desde la admisión. Más aún, estudios recientes advierten que la complejidad clínica relevante para predecir desviaciones importantes en la estancia no siempre se encuentra registrada en las primeras horas del episodio hospitalario (14,47). En este contexto, el presente trabajo confirma que la información inicial disponible en urgencia, demográfica, administrativa y signos vitales, resulta insuficiente para capturar la totalidad de determinantes que tienden a manifestarse durante la evolución hospitalaria.

El análisis descriptivo realizado sobre 13.760 egresos reveló un conjunto coherente de factores asociados a estancias prolongadas: mayor edad, nivel crítico de *triage*, parámetros clínicos iniciales alterados y residencia fuera de la región metropolitana. Estos hallazgos son consistentes con revisiones internacionales que identifican la gravedad inicial, fragilidad, comorbilidad y complejidad diagnóstica como predictores significativos de *outliers* (12,14). Asimismo, la concentración de *outliers* en patologías como sepsis e insuficiencia respiratoria se alinea con la evidencia que muestra que los GRD de naturaleza infecciosa o respiratoria son particularmente propensos a generar desviaciones extensas respecto a la mediana de estancia (48). El comportamiento observado en los GRD asociados a abscesos periamigdalinos o patologías otorrinolaringológicas complejas también replica lo descrito por Bellanger et al. (2013) (17), quienes mostraron un comportamiento similar en procedimientos de cabeza y cuello.

En cuanto a la capacidad predictiva de los modelos evaluados, los resultados evidencian un desempeño moderado, con AUC-ROC entre 0.56 y 0.66. La superioridad relativa de Naive Bayes y Regresión Logística coincide con estudios que destacan su estabilidad en contextos de alta dimensionalidad y datos desbalanceados (22,38), mientras que el rendimiento más débil de Random Forest, XGBoost, LightGBM y MLP refleja su sensibilidad al desbalance extremo del evento (<5%). Este comportamiento ha sido ampliamente reportado en la literatura, donde se advierte que los modelos basados en árboles y redes neuronales tienden a generar fronteras de decisión distorsionadas cuando la clase minoritaria es escasa, aun cuando se aplican técnicas de sobremuestreo sintético (38,49).

El uso de técnicas como SMOTE, SMOTEENN y ADASYN no logró mejorar el rendimiento predictivo, e incluso produjo ligeras disminuciones del AUC en ciertos modelos. Este fenómeno puede interpretarse a la luz de los estudios comparativos que muestran que, cuando la clase minoritaria es extremadamente heterogénea, como ocurre con los *outliers* superiores, la síntesis artificial de casos puede generar instancias poco realistas que introducen ruido y reducen la capacidad de generalización (47,50). En este sentido, los resultados del presente trabajo coinciden con la evidencia reciente que sugiere que, para fenómenos clínicos de baja prevalencia y perfiles diversos, la generación de datos sintéticos no siempre constituye una estrategia eficaz.

Otro aspecto relevante es el efecto del solapamiento clínico entre *inliers* y *outliers*. A pesar de que varios signos vitales mostraron diferencias estadísticamente significativas, los rangos intercuartílicos se superponen de manera sustancial, lo que impide establecer umbrales discriminantes robustos. Este hallazgo ha sido descrito en múltiples estudios de predicción temprana en EHR, donde la variabilidad fisiológica inicial no ofrece suficiente poder discriminante para anticipar eventos complejos como la estancia prolongada, que dependen de complicaciones posteriores, respuesta al tratamiento y factores organizacionales (22,51).

Desde una perspectiva institucional, este estudio también pone en relieve limitaciones operativas asociadas a la calidad y disponibilidad de los datos. La base GRD se destacó por su completitud y estandarización, coherente con lo reportado en experiencias europeas donde la calidad del CMBD es clave para el análisis de desempeño hospitalario (15,52). En contraste, la base de datos del Servicio de Urgencia mostró limitaciones en la profundidad de las variables clínicas y presencia de valores atípicos, situación también descrita en la literatura sobre el uso de EHR para

modelamiento predictivo inicial (50). La exclusión de los años 2020–2021 debido al impacto epidemiológico de la pandemia refuerza la necesidad de realizar validaciones externas en escenarios más habituales.

La interpretabilidad de los datos en este estudio, según los modelos utilizados para predecir *outliers* superiores, principalmente Regresión Logística y Naive Bayes, demostraron ser especialmente adecuados debido a su alta interpretabilidad. Ambos permiten identificar de manera explícita cómo cada una de las variables clínicas iniciales (*triage\_alto*, *sat\_baja*, *glasgow\_bajo*, *servicio\_riesgo*, *diag\_riesgo*, *diag\_protector* y *servicio\_protector*) contribuye al riesgo de prolongación de la estancia hospitalaria. La Regresión Logística, en particular, ofrece coeficientes de fácil interpretación, lo que facilita justificar decisiones clínicas y respaldar procesos de auditoría institucional. Al compararlos con modelos más complejos, como Random Forest, XGBoost, LightGBM o redes neuronales, se observa que estos últimos no superaron el rendimiento de Naive Bayes (AUC 0.663), probablemente debido al desbalance de clases, al número limitado de predictores relevantes disponibles al ingreso y a la heterogeneidad diagnóstica. Además, estos métodos requieren técnicas adicionales de explicabilidad (como SHAP o LIME), lo que incrementa la complejidad y dificulta su implementación práctica en entornos clínicos donde la transparencia es esencial. En este contexto, los modelos interpretables empleados resultan metodológicamente coherentes con los objetivos del estudio. Aunque el rendimiento predictivo fue moderado, la claridad respecto a los factores asociados al riesgo de convertirse en *outlier* superior constituye un aporte relevante para la gestión hospitalaria basada en datos. Así, estos modelos no solo permiten comprender mejor el fenómeno, sino que también pueden integrarse como herramientas de apoyo para la estratificación temprana del riesgo, la planificación de recursos y la mejora de los procesos clínico-administrativos.

Finalmente, las implicancias clínicas y administrativas de este trabajo son relevantes. Los modelos construidos, aunque no alcanzaron una capacidad predictiva adecuada para uso clínico, sí mostraron una excelente identificación de *inliers*, lo que podría permitir su utilización como herramienta de tamizaje negativo en el ingreso. De esta manera, el sistema podría descartar tempranamente a los pacientes con baja probabilidad de prolongarse, focalizando los recursos analíticos y clínicos en los casos potencialmente complejos. Este enfoque es congruente con propuestas recientes de predicción temprana en sistemas de salud que plantean modelos de estratificación progresiva en lugar de decisiones binarias (20,47).

Las limitaciones del estudio incluyeron el desbalance extremo de la clase *outlier* superior, la heterogeneidad diagnóstica del conjunto GRD, la falta de variables longitudinales y de laboratorio al ingreso, el solapamiento en rangos clínicos entre grupos, y la naturaleza retrospectiva de los datos disponibles.

Futuros trabajos deberían integrar datos longitudinales, notas clínicas, resultados de laboratorio, comorbilidades estructuradas, e incluso modelos de aprendizaje profundo basados en secuencias (RNN, transformers), en línea con los desarrollos recientes de DeepDRG y DRG-LLaMA (20,22). Asimismo, es recomendable avanzar hacia validaciones externas multiservicio y estudiar modelos híbridos que combinen información clínica, administrativa y organizacional.

## 9. CONCLUSIONES

El presente estudio permitió desarrollar un análisis sistemático orientado a la identificación temprana de pacientes con riesgo de estancia hospitalaria prolongada y *outliers* superiores a partir de información disponible en el momento del ingreso por el Servicio de Urgencia del Hospital Clínico de la Universidad de Chile (HCUCH). Mediante la aplicación de técnicas de aprendizaje automático supervisado, se buscó establecer un modelo predictivo capaz de anticipar la evolución hospitalaria y contribuir a la optimización de la gestión clínica y de recursos.

Se evidenció que el sistema GRD, si bien constituye una herramienta para la clasificación retrospectiva de episodios hospitalarios, presenta limitaciones significativas en su capacidad predictiva temprana. Los resultados mostraron que los pacientes definidos como *outliers* superiores representan una proporción reducida (4,6 %) del total de egresos, pero concentran una carga asistencial y económica considerable, con una mediana de 27 días de hospitalización, muy superior a la de los *inliers* (6 días).

El análisis de las variables iniciales permitió identificar servicios clínicos y diagnósticos asociados a un mayor riesgo de estancia hospitalaria prolongada. Los servicios de Unidad de Pacientes Críticos (UPC), Otorrinolaringología y Psiquiatría mostraron una mayor proporción de pacientes clasificados como *outliers* superiores, constituyéndose en áreas de mayor riesgo institucional. En contraste, los servicios de Cirugía, Cardiología y Unidad Coronaria se comportaron como factores protectores, asociados a una menor probabilidad de prolongación de la estancia. En el ámbito diagnóstico, los hallazgos identificados previamente en los resultados muestran que ciertos grupos de patologías se asocian de manera diferencial con el riesgo de ser *outlier* superior. En particular, algunos diagnósticos clínicos demostraron una relación significativa con una mayor probabilidad de ser considerado *outlier* superior, mientras que otros se vincularon con un riesgo sustancialmente menor. Estos hallazgos refuerzan la importancia de la evaluación clínica inicial y la identificación temprana de diagnósticos de riesgo como elementos clave para anticipar la evolución del paciente y optimizar la planificación hospitalaria y la asignación de recursos desde el momento del ingreso.

Los algoritmos Naive Bayes (AUC = 0.663) y Regresión Logística (AUC = 0.662) alcanzaron el mejor rendimiento global, seguidos de XGBoost (AUC = 0.652). Este rendimiento sugiere que, si bien las herramientas de *machine learning* son prometedoras, su eficacia depende de una mayor

calidad, profundidad y balance de los datos clínicos utilizados para el entrenamiento. La baja prevalencia de *outliers* superiores y la heterogeneidad diagnóstica constituyen factores limitantes relevantes que deben abordarse en futuras versiones del modelo. En conjunto, los hallazgos permiten concluir que la predicción precoz de *outliers* superiores es factible, pero requiere la integración de fuentes de datos más amplias, incluyendo variables fisiológicas, evolutivas y contextuales, así como el uso de técnicas avanzadas de balanceo de clases y optimización de hiperparámetros.

Finalmente, este estudio sienta las bases para el desarrollo de un sistema predictivo de apoyo a la decisión clínica en el contexto hospitalario, orientado a optimizar la asignación de recursos y mejorar la eficiencia operacional. Aunque los modelos evaluados no lograron predecir con alta exactitud a los pacientes con estancias prolongadas (*outliers* superiores), sí demostraron una capacidad notablemente superior para identificar con precisión a aquellos cuya evolución se mantendría dentro de los rangos esperados (*inliers*). Este hallazgo, lejos de ser una limitación, constituye un aporte relevante: disponer de una herramienta confiable para reconocer tempranamente a los pacientes con bajo riesgo de prolongación permite focalizar los esfuerzos clínicos y administrativos en el subgrupo más complejo y potencialmente problemático.

La elevada certidumbre en la predicción de *inliers* ofrece un valor estratégico en el Servicio de Urgencia, al posibilitar un tamizaje inicial que contribuya a una planificación más eficiente de camas, recursos críticos y tiempos de hospitalización. En un entorno donde la disponibilidad de recursos es limitada, contar con un modelo que discrimine con claridad a los pacientes de bajo riesgo permite redirigir la atención hacia aquellos con mayor probabilidad de evolucionar desfavorablemente, anticipando necesidades y facilitando una gestión más proactiva.

En perspectiva, la integración de este enfoque predictivo en los procesos de admisión hospitalaria podría constituir un instrumento relevante para la gestión clínica y la planificación institucional. Si se complementa en el futuro con variables clínicas más ricas, información longitudinal y modelos híbridos, este sistema podría evolucionar hacia una herramienta robusta y aplicable en tiempo real, alineada con los principios de eficiencia, seguridad del paciente y atención centrada en la persona.

## 10. PERSPECTIVAS FUTURAS

**Refinamiento metodológico y expansión:** Una de las principales proyecciones de este trabajo consiste en profundizar y ampliar el modelo predictivo mediante la incorporación de nuevas fuentes de datos clínicos. En una futura fase, resulta fundamental integrar variables derivadas de exámenes de laboratorio, tales como hemograma, marcadores inflamatorios (PCR, procalcitonina), función renal, electrolitos y gases arteriales. Estas variables poseen un alto poder discriminativo en cuadros clínicos complejos y podrían mejorar sustancialmente la capacidad del modelo para identificar precozmente a los *outliers* superiores. Asimismo, la integración de variables temporales, como la evolución de signos vitales, registros seriados de laboratorio o intervenciones diagnósticas durante las primeras horas de hospitalización, permitiría capturar la dinámica clínica del paciente, avanzando hacia modelos temporales más robustos, como LSTM o modelos secuenciales.

Un segundo aspecto dentro del refinamiento metodológico se relaciona con la optimización del proceso de selección de variables. Si bien en esta tesis se utilizó un enfoque basado en *Forward Selection* y criterios de significancia y rendimiento, futuras investigaciones podrían explorar técnicas más avanzadas, como selección basada en regularización (LASSO, *Elastic Net*), algoritmos genéticos o métodos de *embedded feature selection* propios de modelos de ensamblaje. Esto permitiría identificar con mayor precisión cuáles variables, demográficas, administrativas o clínicas, aportan efectivamente al rendimiento del modelo e incorporar únicamente aquellas que maximicen la generalización.

Adicionalmente, una línea relevante es ampliar el período de predicción y la ventana de análisis, evaluando si el modelo puede anticiparse no solo al ingreso, sino también durante las primeras 6, 12 o 24 horas de estancia hospitalaria. Esto abriría la posibilidad de evaluar modelos de predicción dinámica, donde el riesgo de convertirse en outlier superior puede actualizarse conforme avanza la hospitalización, enriqueciendo la utilidad clínica del modelo. Del mismo modo, incorporar técnicas de Procesamiento de Lenguaje Natural (PLN) para analizar texto libre proveniente de notas médicas, epicrisis o motivos de consulta permitiría capturar información cualitativa que actualmente no es utilizada y que podría ser altamente predictiva

**Integración de un modelo en la práctica hospitalaria:** Un paso natural tras los resultados obtenidos es avanzar hacia un proceso de implementación operativa del modelo en el entorno clínico real del HCUCH. Esto implica desarrollar un prototipo de herramienta informática o panel de riesgo integrado a los sistemas de información hospitalaria que permita alertar tempranamente sobre pacientes con alta probabilidad de transformarse en *outliers* superiores. La validación del modelo en condiciones reales requerirá un proceso de co-diseño con equipos clínicos, codificadores GRD y unidades de gestión hospitalaria, para asegurar que el sistema sea interpretativo, clínicamente útil y operacionalmente factible. Asimismo, futuras investigaciones deberán evaluar el impacto clínico y organizacional del modelo, midiendo indicadores como disminución de estancias prolongadas, uso eficiente de camas, reasignación de recursos o capacidad de anticipación de equipos clínicos.

**Análisis de *inliers* superiores:** Finalmente, un ámbito de investigación complementario y altamente relevante para este proyecto es el análisis de los *inliers* superiores, definidos como aquellos pacientes cuya estancia hospitalaria resulta anormalmente corta respecto al tiempo esperado para su GRD. Aunque con frecuencia se asocian a episodios clínicamente más simples o a una resolución asistencial altamente eficiente, su estudio también puede revelar potenciales inconsistencias, variabilidad injustificada o problemas en el proceso de atención.

Desde la perspectiva de esta tesis, centrada en la detección precoz de *outliers* superiores, incorporar el análisis de los *inliers* superiores permitiría comprender mejor la distribución de la estancia hospitalaria. Además, el estudio de este grupo permitiría identificar patrones de alta eficiencia asistencial que podrían servir como referencia para mejorar la gestión hospitalaria, optimizar flujos, y promover buenas prácticas.

Incorporar el análisis de los *inliers* superiores en investigaciones futuras contribuiría, por tanto, a obtener una visión más integral del comportamiento de los tiempos de hospitalización en el HCUCH. De este modo, sería posible complementar la predicción de *outliers* superiores con una evaluación más amplia de la variabilidad asistencial y de los factores que influyen tanto en las estancias excesivamente prolongadas como en las anormalmente cortas. Esto permitiría reforzar el aporte de los GRD como herramienta de monitoreo, auditoría y mejora continua en los procesos hospitalarios.



## 11. BIBLIOGRAFÍA

1. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. *Med Care.* febrero de 1980;18(2 Suppl):iii, 1-53.
2. Zou K, Li HY, Zhou D, Liao ZJ. The effects of diagnosis-related groups payment on hospital healthcare in China: a systematic review. *BMC Health Serv Res.* diciembre de 2020;20(1):112.
3. M. MZ. IMPORTANCIA DEL SISTEMA GRD PARA ALCANZAR LA EFICIENCIA HOSPITALARIA. *Revista Médica Clínica Las Condes.* mayo de 2018;29(3):347-52.
4. Águila R A, Muñoz D MA, Sepúlveda S. V. Experiencia en el desarrollo e implementación de la metodología de grupos relacionados por diagnóstico en un hospital universitario chileno. Evaluación a diez años de funcionamiento. *Rev méd Chile.* diciembre de 2019;147(12):1518-26.
5. Liu X, Fang C, Wu C, Yu J, Zhao Q. DRG grouping by machine learning: from expert-oriented to data-based method. *BMC Med Inform Decis Mak.* diciembre de 2021;21(1):312.
6. Manterola C, García MF, Muñoz G, Holguín JP. CONJUNTO MÍNIMO DE DATOS. SU UTILIDAD EN CIRUGÍA. *Rev Cirugia [Internet].* 13 de junio de 2022 [citado 22 de julio de 2023];74(3). Disponible en: <https://www.revistacirugia.cl/index.php/revistacirugia/article/view/1402>
7. Decreto-Exento-643-Sustituye-Norma-Técnica-sobre-Estándares-de-Información-de-Salud-Actualizada-a-Dic-2016.pdf.
8. Mathauer I, Wittenbecher F. DRG-based payment systems in low- and middle-income countries: Implementation experiences and challenges.
9. NMIRv N. Norma MINSAL 2018-2019. Grupos Relacionados por Diagnóstico Refinados internacionales (GRD IR).
10. Chok L, Bachli EB, Steiger P, Bettex D, Cottini SR, Keller E, et al. Effect of diagnosis related groups implementation on the intensive care unit of a Swiss tertiary hospital: a cohort study. *BMC Health Serv Res.* diciembre de 2018;18(1):84.
11. Documento conductor de Reportes Mensuales de Producción Hospitalaria e Indicadores GRD Minsal.
12. Jentzsch T, Seifert B, Neuhaus V, Moos RM. Predictors for shorter and longer length of hospital stay outliers: a retrospective case-control study of 8247 patients at a university

- hospital trauma department. Swiss Med Wkly [Internet]. 22 de agosto de 2018 [citado 22 de octubre de 2024]; Disponible en: <https://smw.ch/index.php/smw/article/view/2502>
13. Yigit P, Aydin S, Güler H. Analyzing Hospital High Length of Stay Outliers in Turkey. *Journal of Basic and Clinical Health Sciences*. 29 de septiembre de 2022;6(3):903-12.
  14. Freitas A, Silva-Costa T, Lopes F, Garcia-Lema I, Teixeira-Pinto A, Brazdil P, et al. Factors influencing hospital high length of stay outliers. *BMC Health Serv Res*. diciembre de 2012;12(1):265.
  15. Peltola M, Quentin W, on behalf of the EuroDRG Group. Diagnosis-Related Groups for Stroke in Europe: Patient Classification and Hospital Reimbursement in 11 Countries. *Cerebrovasc Dis*. 2013;35(2):113-23.
  16. Flohé S, Buschmann C, Nabring J, Merguet P, Luetkes P, Lefering R, et al. Polytraumadefinition im G-DRG-System 2006: Bis zu 30% „Fehlgruppierungen“. *Unfallchirurg*. julio de 2007;110(7):651-8.
  17. Bellanger MM, Quentin W, Tan SS. Childbirth and Diagnosis Related Groups (DRGs): patient classification and hospital reimbursement in 11 European countries. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. mayo de 2013;168(1):12-9.
  18. Barrientos Vega R. Nuestra experiencia con los grupos relacionados por el diagnóstico en una unidad de cuidados intensivos. *Medicina Intensiva*. enero de 2003;27(6):391-8.
  19. Chen Y jing, Zhang X yu, Yan J qi, Xue-Tang, Qian M cen, Ying X hua. Impact of Diagnosis-Related Groups on Inpatient Quality of Health Care: A Systematic Review and Meta-Analysis. *INQUIRY*. enero de 2023;60:004695802311670.
  20. Wang H, Gao C, Dantona C, Hull B, Sun J. DRG-LLaMA : tuning LLaMA model to predict diagnosis-related group for hospitalized patients. *npj Digit Med*. 22 de enero de 2024;7(1):16.
  21. Liu J, Capurro D, Nguyen A, Verspoor K. Early prediction of diagnostic-related groups and estimation of hospital cost by processing clinical notes. *npj Digit Med*. 1 de julio de 2021;4(1):103.
  22. Islam MdM, Li GH, Poly TN, Li YC (Jack). DeepDRG: Performance of Artificial Intelligence Model for Real-Time Prediction of Diagnosis-Related Groups. *Healthcare*. 25 de noviembre de 2021;9(12):1632.
  23. Mit Critical Data. Secondary Analysis of Electronic Health Records [Internet]. Cham: Springer International Publishing; 2016 [citado 5 de diciembre de 2025]. Disponible en: <http://link.springer.com/10.1007/978-3-319-43742-2>
  24. Review\_of\_Data\_Preprocessing\_Techniques.

25. Huang X, Jin G, Ruan W. Machine Learning Safety [Internet]. Singapore: Springer Nature Singapore; 2023 [citado 16 de noviembre de 2025]. (Artificial Intelligence: Foundations, Theory, and Algorithms). Disponible en: <https://link.springer.com/10.1007/978-981-19-6814-3>
26. Dietrich F, Schilders W. Scientific machine learning. Math Semesterber. octubre de 2025;72(2):89-115.
27. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3. Aufl. Hoboken, N.J: Wiley; 2013. 500 p. (Wiley Series in Probability and Statistics).
28. Random Forests.
29. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. En: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. San Francisco California USA: ACM; 2016 [citado 16 de noviembre de 2025]. p. 785-94. Disponible en: <https://dl.acm.org/doi/10.1145/2939672.2939785>
30. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree.
31. Cortes C, Vapnik V. Support-vector networks. Mach Learn. septiembre de 1995;20(3):273-97.
32. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. BMJ. 3 de septiembre de 2024;386:e078276.
33. Rani V, Saxena B, Atale N, editores. Advancing Biotechnology: From Science to Therapeutics and Informatics: Technological Advancements in Biosciences and Biotechnology [Internet]. Cham: Springer Nature Switzerland; 2025 [citado 16 de noviembre de 2025]. (Advances in Science, Technology & Innovation). Disponible en: <https://link.springer.com/10.1007/978-3-031-80973-6>
34. Singh P, Mishra AR, Garg P, editores. Data Analytics and Machine Learning: Navigating the Big Data Landscape [Internet]. Singapore: Springer Nature Singapore; 2024 [citado 16 de noviembre de 2025]. (Studies in Big Data; vol. 145). Disponible en: <https://link.springer.com/10.1007/978-981-97-0448-4>
35. Kazi JU. Python Essentials for Biomedical Data Analysis: An Introductory Textbook [Internet]. Cham: Springer Nature Switzerland; 2025 [citado 16 de noviembre de 2025]. Disponible en: <https://link.springer.com/10.1007/978-3-031-85600-6>
36. Yang C, Fridgerisson EA, Kors JA, Reps JM, Rijnbeek PR. Impact of random oversampling and random undersampling on the performance of prediction models developed using observational health data. J Big Data. 3 de enero de 2024;11(1):7.

37. Husain G, Nasef D, Jose R, Mayer J, Bekbolatova M, Devine T, et al. SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models. *Algorithms*. 10 de enero de 2025;18(1):37.
38. Haibo He, Yang Bai, Garcia EA, Shutao Li. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. En: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) [Internet]. Hong Kong, China: IEEE; 2008 [citado 18 de noviembre de 2025]. p. 1322-8. Disponible en: <http://ieeexplore.ieee.org/document/4633969/>
39. Yacoub R, Axman D. Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. En: Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems [Internet]. Online: Association for Computational Linguistics; 2020 [citado 18 de noviembre de 2025]. p. 79-91. Disponible en: <https://www.aclweb.org/anthology/2020.eval4nlp-1.9>
40. Jaakkola R, Janhunen T, Kuusisto A, Rankooh MF, Vilander M. Interpretable classifiers for tabular data via feature selection and discretization.
41. Ghosh B, Malioutov D, Meel KS. Efficient Learning of Interpretable Classification Rules. *jair*. 30 de agosto de 2022;74:1823-63.
42. Krishnamoorthy S. Interpretable Classifier Models for Decision Support Using High Utility Gain Patterns. *IEEE Access*. 2024;12:126088-107.
43. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*. 1 de enero de 2013;20(1):144-51.
44. Abughazala M, Ibiyo M, Muccini H, Sharaf M. Quality by Prompt: LLM-Powered Transformation of Data Quality Requirements Into Great Expectations. En: Taibi D, Smite D, editores. *Software Engineering and Advanced Applications* [Internet]. Cham: Springer Nature Switzerland; 2026 [citado 15 de noviembre de 2025]. p. 130-47. (Lecture Notes in Computer Science; vol. 16081). Disponible en: [https://link.springer.com/10.1007/978-3-032-04190-6\\_9](https://link.springer.com/10.1007/978-3-032-04190-6_9)
45. Fetter RB. Diagnosis Related Groups: Understanding Hospital Performance. *Interfaces*. febrero de 1991;21(1):6-26.
46. Busse R, Geissler A, Aaviksoo A, Cots F, Hakkinen U, Kobel C, et al. Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals? *BMJ*. 7 de junio de 2013;346(jun07 3):f3197-f3197.

47. Stone K, Zwigelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: Towards a unified framework. Yoon D, editor. PLOS Digit Health. 14 de abril de 2022;1(4):e0000017.
48. Lin S, Rouse P, Wang YM, Zhang F. A Statistical Model to Detect DRG Outliers. IEEE Access. 2022;10:28717-24.
49. Brandt J, Lanzén E. A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification.
50. Nallabasannagari AR, Reddiboina M, Seltzer R, Zeffiro T, Sharma A, Bhandari M. All Data Inclusive, Deep Learning Models to Predict Critical Events in the Medical Information Mart for Intensive Care III Database (MIMIC III) [Internet]. arXiv; 2020 [citado 15 de noviembre de 2025]. Disponible en: <http://arxiv.org/abs/2009.01366>
51. Meng Z, Zou K, Song S, Wu H, Han Y. Associations of Chinese diagnosis-related group systems with inpatient expenditures for older people with hip fracture. BMC Geriatr. diciembre de 2022;22(1):169.
52. Schuster M, Kuntz L, Hermening D, Bauer M, Abel K, Goetz AE. Die Nutzung der Erlösdaten der „DRGs“ für ein externes Benchmarking der anästhesiologischen und intensivmedizinischen Leistungserbringung. Anaesthesist. enero de 2006;55(1):26-32.

## 12. ANEXOS

Anexo N° 01: Categorías diagnósticas mayores (CDM)

CDM	Descripción
1	Enfermedades y trastornos del sistema nervioso
2	Enfermedades y trastornos del ojo
3	Enfermedades y trastornos de oído, nariz, boca y garganta
4	Enfermedades y trastornos del aparato respiratorio
5	Enfermedades y trastornos del aparato circulatorio
6	Enfermedades y trastornos del aparato digestivo
7	Enfermedades y trastornos de hígado, sistema biliar y páncreas
8	Enfermedades y trastornos del sistema músculo-esquelético y tejido conectivo
9	Enfermedades y trastornos de piel, tejido subcutáneo y mama
10	Enfermedades y trastornos del sistema endocrino, nutricional y metabólico
11	Enfermedades y trastornos del aparato urinario
12	Enfermedades y trastornos del aparato reproductor masculino
13	Enfermedades y trastornos del aparato reproductor femenino
14	Parto
15	Recién nacidos y otros neonatos
16	Enfermedades y trastornos de sangre, órganos hematopoyéticos y del sistema inmunológico
17	Enfermedades y trastornos mieloproliferativos y neoplasias mal diferenciadas
18	Enfermedades infecciosas y parasitarias sistémicas o de sitios no especificados
19	Enfermedades y trastornos mentales
20	Abuso y dependencia de drogas y alcohol
21	Lesiones, envenenamientos y efectos tóxicos de fármacos
22	Factores que influyen en el estado de salud y en el contacto con los servicios sanitarios
23	Visitas médicas ambulatorias

Anexo N° 02: Tipos de IR-GRD

Tipo	Descripción
1	Procedimientos en pacientes hospitalizados
2	Procedimientos mayores en pacientes ambulatorios
3	Procedimiento significativo ambulatorio
4	Médicos para pacientes hospitalizados
5	Médicos para pacientes ambulatorios
6	Partos hospitalizados
7	Partos ambulatorios
8	Recién nacidos hospitalizados
9	Recién nacidos ambulatorios
0	Error

## CERTIFICADO

El Comité Ético Científico y de Investigación (CEIC) en Seres Humanos del Hospital Clínico Universidad de Chile, ha recibido el formulario de Solicitud para revisión de:  
**Base de datos GRD obtenidos de los egresos hospitalarios HCUCH, desde el 2001.**

Cuyo título de trabajo es:

**“Análisis inteligente de la data de grupos relacionados a diagnóstico (GRD) de pacientes egresados de unidades críticas de un hospital universitario, como herramienta para optimizar la gestión clínica y financiera”.**

**La Dra. Nivia Estuardo Agurto, Jefa de la Unidad de Pacientes Críticos,** afirma estar en conocimiento del trabajo mencionado y autoriza su realización.

Se otorga el certificado al investigador:

**Dr. Jorge Montoya Zumaeta, de la Unidad de Pacientes Críticos.**

El estudio fue analizado en conformidad, en sesión del 14 de junio de 2023, y no existen reparos éticos con su aprobación.



14-06-2023

Dra. Lucía Teresa Massardo Vega  
Presidenta del Comité Ético Científico  
Hospital Clínico de la Universidad de Chile