



Premio de Investigación SCHOT 2020: desarrollo y validación de un modelo multivariable de predicción de estadía hospitalaria en pacientes mayores de 65 años sometidos a artroplastia total de cadera electiva en Chile utilizando aprendizaje de máquinas

2020 SCHOT Research Award: Development and Validation of a Multivariable Prediction Model of Hospital Stay in Elderly Chilean Patients Undergoing Elective Total Hip Arthroplasty Using Machine Learning

Claudio Díaz-Ledezma^{1,2} David Díaz-Solís³ Raúl Muñoz-Reyes⁴ Jonathan Torres Castro^{5,6}

¹Unidad de Cirugía Ortopédica y Traumatología, Hospital El Carmen Dr. Luis Valentin Ferrada, Santiago, Chile

²Departamento de Ortopedia y Traumatología, Clínica Las Condes, Santiago, Chile

³Departamento de Administración, Facultad de Economía y Negocios, Universidad de Chile, Santiago, Chile

⁴Data scientist, independent researcher, Santiago, Chile

⁵Equipo de Cirugía de Cadera, Clínica RedSalud Santiago, Santiago, Chile

⁶Equipo de Cirugía de Cadera, Instituto Traumatológico de Santiago, Santiago, Chile

Dirección para correspondencia Claudio Díaz Ledezma, MD, Av. Rinconada 1.201, Oficina 28, 5to piso, Maipú, Santiago, Chile (e-mail: claudioadiaz@gmail.com).

Rev Chil Ortop Traumatol 2021;62(3):e180–e192.

Resumen

Palabras Clave

- ▶ estadía hospitalaria
- ▶ aprendizaje de máquinas
- ▶ artroplastia total de cadera

Introducción La predicción de la estadía hospitalaria luego de una artroplastia total de cadera (ATC) electiva es crucial en la evaluación perioperatoria de los pacientes, con un rol determinante desde el punto de vista operacional y económico. Internacionalmente, se han empleado macrodatos (*big data*, en inglés) e inteligencia artificial para llevar a cabo evaluaciones pronósticas de este tipo. El objetivo del presente estudio es desarrollar y validar, con el empleo del aprendizaje de máquinas (*machine learning*, en inglés), una herramienta capaz de predecir la estadía hospitalaria de pacientes chilenos mayores de 65 años sometidos a ATC por artrosis.

recibido

18 de marzo de 2021

aceptado

06 de agosto de 2021

DOI <https://doi.org/10.1055/s-0041-1740232>.
ISSN 0716-4548.

© 2021. Sociedad Chilena de Ortopedia y Traumatología. All rights reserved.

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Thieme Revinter Publicações Ltda., Rua do Matoso 170, Rio de Janeiro, RJ, CEP 20270-135, Brazil

Material y Métodos Empleando los registros electrónicos de egresos hospitalarios anonimizados del Departamento de Estadísticas e Información de Salud (DEIS), se obtuvieron los datos de 8.970 egresos hospitalarios de pacientes sometidos a ATC por artrosis entre los años 2016 y 2018. En total, 15 variables disponibles en el DEIS, además del porcentaje de pobreza de la comuna de origen del paciente, fueron incluidos para predecir la probabilidad de que un paciente presentara una estadía acortada (< 3 días) o prolongada (> 3 días) luego de la cirugía. Utilizando técnicas de aprendizaje de máquinas, 8 algoritmos de predicción fueron entrenados con el 80% de la muestra. El 20% restante se empleó para validar las capacidades predictivas de los modelos creados a partir de los algoritmos. La métrica de optimización se evaluó y ordenó en un *ranking* utilizando el área bajo la curva de característica operativa del receptor (*area under the receiver operating characteristic curve*, AUC-ROC, en inglés), que corresponde a cuan bien un modelo puede distinguir entre dos grupos.

Resultados El algoritmo XGBoost obtuvo el mejor desempeño, con una AUC-ROC promedio de 0,86 (desviación estándar [DE]: 0,0087). En segundo lugar, observamos que el algoritmo lineal de máquina de vector de soporte (*support vector machine*, SVM, en inglés) obtuvo una AUC-ROC de 0,85 (DE: 0,0086). La importancia relativa de las variables explicativas demostró que la región de residencia, el servicio de salud, el establecimiento de salud donde se operó el paciente, y la modalidad de atención son las variables que más determinan el tiempo de estadía de un paciente.

Discusión El presente estudio desarrolló algoritmos de aprendizaje de máquinas basados en macrodatos chilenos de libre acceso, y logró desarrollar y validar una herramienta que demuestra una adecuada capacidad discriminatoria para predecir la probabilidad de estadía hospitalaria acortada *versus* prolongada en adultos mayores sometidos a ATC por artrosis.

Conclusión Los algoritmos creados a través del empleo del aprendizaje de máquinas permiten predecir la estadía hospitalaria en pacientes chilenos operado de artroplastia total de cadera electiva.

Abstract

Introduction The prediction of the length of hospital stay after elective total hip arthroplasty (THA) is crucial in the perioperative evaluation of the patients, and it plays a decisive role from the operational and economic point of view. Internationally, big data and artificial intelligence have been used to perform prognostic evaluations of this type. The present study aims to develop and validate, through the use of artificial intelligence (machine learning), a tool capable of predicting the hospital stay of patients over 65 years of age undergoing THA for osteoarthritis.

Material and Methods Using the electronic records of hospital discharges de-identified from the Department of Health Statistics and Information (Departamento de Estadísticas e Información de Salud, DEIS, in Spanish), the data of 8,970 hospital discharges of patients who had undergone THA for osteoarthritis between 2016 and 2018 were obtained. A total of 15 variables available in the DEIS registry, in addition to the poverty rate in the patient's borough of origin were included to predict the probability that a patient would have a shortened (< 3 days) or prolonged (> 3 days) stay after surgery. By using machine learning techniques, 8 prediction algorithms were trained with 80% of the sample. The remaining 20% was used to validate the predictive capabilities of the models created from the algorithms. The optimization metric was evaluated and ranked using the area under the receiver operating characteristic curve (AUC-ROC), which corresponds to how well a model can distinguish between two groups.

Results The XGBoost algorithm had the best performance, with an average AUC-ROC of 0.86 (standard deviation [SD]: 0.0087). Secondly, we observed that the linear

support vector machine (SVM) algorithm obtained an AUC-ROC of 0.85 (SD: 0.0086). The relative importance of the explanatory variables showed that the region of residence, the administrative health service, the hospital where the patient was operated on, and the care modality are the variables that most determine the length of stay.

Discussion The present study developed machine learning algorithms based on free-access Chilean big data, which helped create and validate a tool that demonstrates an adequate discriminatory capacity to predict shortened versus prolonged hospital stay in elderly patients undergoing elective THA.

Conclusion The algorithms created through the use of machine learning allow to predict the hospital stay in Chilean patients undergoing elective total hip arthroplasty.

Keywords

- ▶ hospital stay
- ▶ machine learning
- ▶ total hip arthroplasty

Introducción

En Chile, la artroplastia total de cadera (ATC) para el tratamiento de artrosis severa está garantizada por ley para pacientes mayores de 65 años.¹ Sin embargo, poco se conoce de los resultados de la ATC en este grupo particular de pacientes, pues no existe (por lo que sabemos) ninguna publicación científica nacional que aborde el tema de la estadía hospitalaria, la cual tiene un papel protagónico en la era de la artroplastia basada en valor.

En el mundo y particularmente en EEUU, se ha observado una baja sostenida en la estadía hospitalaria de los pacientes tras ATC, sin aumentar los riesgos.² Inclusive, se ha probado que la modalidad ambulatoria puede ser exitosa en un grupo selecto de pacientes.^{3,4} La estadía hospitalaria para pacientes mayores de 65 años en EEUU (2015-2016) fue en promedio de 1,8 días.⁵ En Chile, estos datos no han sido publicados.

Varias tácticas se pueden utilizar para disminuir la estadía hospitalaria en casos de ATC, entre ellas los protocolos estandarizados de manejo,^{6,7} y otras, que van de la mano con la predicción de las potenciales complicaciones perioperatorias.^{8,9} Entre los desafíos de la ATC en nuestro país, hemos descrito la relevancia de mantener nuestro enfoque actualizado y con los mismos estándares que los de los países líderes en el tema.¹⁰

Conforme avanzamos en la crisis global de la pandemia COVID-19, se ha enfatizado el hecho de que las cirugías electivas se realicen con un alta hospitalaria lo más precoz posible, sin comprometer la seguridad del paciente.^{11,12} Entre las tareas relevantes que tenemos los cirujanos está el intentar predecir la posibilidad de complicaciones y la duración de la estadía hospitalaria de nuestros pacientes.

El aprendizaje de máquinas (*machine learning*, en inglés) es una rama de la inteligencia artificial¹³ entendida como el estudio de la forma en que algoritmos informáticos (es decir, las máquinas) pueden “aprender” relaciones o patrones complejos a partir de datos empíricos, y, por lo tanto, producir modelos matemáticos que vinculan un número grande de covariables a alguna variable que sea objetivo de interés.¹⁴

En el campo médico, entre otras aplicaciones, esto significa poder predecir, a partir de datos extraídos de registros electrónicos especializados, puntajes de riesgo (en forma de regresión y de pronóstico) para ayudar a los médicos a tomar decisiones más eficientes y precisas; por lo tanto, el aprendizaje de máquinas puede ser una herramienta de apoyo en las decisiones clínicas. Específicamente en artroplastia, estudios¹⁵⁻¹⁷ involucrando esta tecnología han ganado momento, y proponen asistencia para resolver complejos problemas que enfrentamos en nuestra práctica.¹⁸

Nuestra hipótesis es la de que, con el proceso de aprendizaje de máquinas, se puede predecir la estadía hospitalaria en pacientes sometidos a ATC, con un doble propósito en la actividad clínica: 1) ayudar a mejorar al grupo con alta probabilidad de estadía corta, disminuyendo aún más su estadía; e 2) identificar al grupo de baja probabilidad de estadía corta, para mejorar su cuidado perioperatorio y finalmente llevarlo al grupo de estadía corta de manera segura.

El objetivo del presente estudio es desarrollar y validar, por el empleo del aprendizaje de máquinas, una herramienta que sea capaz de predecir la estadía hospitalaria de pacientes mayores de 65 años sometidos a ATC por artrosis.

Material y Métodos

Fondos Relacionados

El presente trabajo fue desarrollado gracias a los fondos de investigación de la Sociedad Chilena de Ortopedia y Traumatología, por medio de su concurso de investigación 2020.

Fuente de Datos y Población de Estudio

El presente corresponde a un estudio de registro. Desde la página *web* del Departamento de Estadísticas e Información en Salud (DEIS) del Ministerio de Salud de Chile,¹⁹ se recogieron las bases de datos de egresos hospitalarios de los años 2016, 2017, y 2018. Cada una de estas bases de datos contiene los registros anonimizados de todos los egresos hospitalarios tanto de centros públicos como privados de

nuestro país, incluyendo 39 columnas con datos relacionados a cada uno de los egresos hospitalarios individualizados. Cada uno de estos datos contiene características demográficas, del centro hospitalario, de la condición de egreso, del diagnóstico etc. En el período estudiado, se recogieron los datos de 4.944.017 egresos hospitalarios. Considerando las 39 columnas antes mencionadas, el volumen total de datos individuales a discriminar y evaluar fue de 192.816.663.

Considerando que los datos de cada caso en particular son no identificados y provenientes de una base de datos pública (la identificación es un código alfanumérico sin datos que identifiquen cada paciente), el presente estudio no requiere autorización de comité de ética.

A partir de la fuente de datos primaria, se creó una base de datos derivada, que incluía solamente a pacientes de edad ≥ 65 años sometidos a artroplastia (o endoprótesis) total de cadera por artrosis, casos que están cubiertos bajo las Garantías Explícitas en Salud.¹ Estos casos fueron seleccionados mediante los códigos del Fondo Nacional de Salud (Fonasa) 2104129 (*Endoprótesis total de cadera, no incluye prótesis*) y 2104229 (*Endoprótesis total de cadera, incluye prótesis*), que correspondan al diagnóstico M16 (coxartrosis) de la Clasificación Internacional de Enfermedades, 10.^a revisión (CIE-10), con todas sus clasificaciones secundarias. Se incluyeron pacientes de todas las previsiones operados a lo largo de Chile en el período 2016-2018. Se excluyeron procedimientos codificados como 2104129 y 2104229 realizados por diagnóstico de fractura del fémur proximal (diagnóstico S72 en el CIE-10) y los casos que egresaron del hospital categorizados como “fallecidos”. La muestra incluyó todos los casos registrados en nuestro país para el período indicado.

Resultado Clínicamente Relevante (Variable a Predecir)

De acuerdo a la literatura,²⁰ una estadía de más de tres días puede considerarse prolongada en el contexto de una ATC electiva. En nuestro estudio, se definirá *estadía corta* como aquella menor o igual a tres días, y *estadía prolongada*, como aquella mayor a tres días, considerando que, para el período estudiado, la experiencia en cirugía de ATC ambulatoria era limitada sólo a algunos grupos en nuestro país.⁴

Se realizó una predicción de la estadía hospitalaria como una variable binaria, descrita en función de dos clases a partir de los días de hospitalización. Así, la variable a modelar toma dos posibles valores: “estadía corta” o “estadía prolongada”.

VARIABLES PREDICTORAS

De las 39 variables individuales para cada 1 de los egresos hospitalarios del DEIS correspondientes a la población de estudio, se eligieron 21 (– **Tabla 1**), consideradas relevantes por el grupo de autores al momento de procesar los datos. Los registros de datos estaban completos para cada una de las variables. De ellas, 16 se utilizaron al momento de realizar un proceso predictivo del alta hospitalaria. Además, se incluyó la variable “porcentaje de pobreza comunal” extraída de la base de datos del Ministerio de Desarrollo Social.²¹ No hubo datos

perdidos en el registro utilizado, por lo que no fue necesario usar técnicas de imputación.²² Es importante destacar que la base de datos del DEIS contiene variables recogidas con propósitos epidemiológicos, y no captura suficientes datos a nivel individual de los pacientes, excluyéndose de este modelo variables como comorbilidades, funcionalidad, y detalles quirúrgicos que ciertamente podrían influenciar la estadía hospitalaria.

Preparación de los Datos (Balanceo de Muestra)

Para el correcto procesamiento de las variables nominales, se procedió a transformarlas usando codificación *one-hot*, es decir, se generaron múltiples columnas dicotómicas que representaban la existencia o no de una característica en particular para cada egreso hospitalario en específico. En términos del procesamiento de las variables continuas, se estandarizó su escala en el rango entre 0 y 1, correspondiendo 0 al mínimo valor en los datos originales, y 1, al máximo para cada uno de ellos. Además, dado que existe una mayor proporción de casos con tres o más días, fue necesario equilibrar la muestra de entrenamiento²³ siguiendo un procedimiento de sobremuestreo de la clase subrepresentada.²⁴

Entrenamiento y Prueba de los Algoritmos de Clasificación

Para el estudio, se probaron diferentes algoritmos y configuraciones de hiperparámetros de estos, disponibles en librerías de código computacional para el lenguaje de programación Python. En particular, se probaron siete algoritmos disponibles en el paquete de *sklearn* (regresión logística, clasificador de árbol de decisión, máquina de vector de soporte lineal, clasificador bayesiano ingenuo, bosques aleatorios, *adaboost*, y perceptrón multicapa). Si bien una descripción detallada del funcionamiento de cada algoritmo está fuera del alcance de los objetivos de este artículo, la intuición detrás de dicha selección hace referencia a la compensación entre poder predictivo y la posible capacidad de interpretación y transparencia de los modelos creados (haciendo que la evaluación de los predictores del modelo no estén bajo la influencia de los autores, una vez que han sido integrados al proceso). En la literatura de aprendizaje de máquinas, es habitual agrupar a los algoritmos si, como estrategia fundamental de modelamiento, ocupan sistemas de ecuaciones matemáticas, o si generan reglas de decisión computacionales, tendiendo éstas últimas a ser más sencillas de interpretar. Los modelos más avanzados, como bosques aleatorios o perceptrón multicapa (un tipo de redes neuronales artificiales), pueden contener miles de reglas de decisión o ecuaciones matemáticas, que tienen por consecuencia potencialmente millones de parámetros a estimar e interpretar. Así, los algoritmos de regresión logística, máquina de vector de soporte, clasificador bayesiano ingenuo, y perceptrón multicapa son basados en sistemas de ecuaciones matemáticas. Por otra parte, los algoritmos de clasificador de árbol de decisión, bosques aleatorios, y *adaboost* generan un conjunto de reglas de decisión computacionales.

Tabla 1 Variables disponibles en la base de datos del Departamento de Estadísticas e Información de Salud (DEIS)

Ítem de la base de datos de egresos hospitalarios del DEIS				
N°	Nombre de la variable	Descripción	Tipo de dato	Utilizado en Modelo
1	ID_PACIENTE	Identificador único y anónimo del paciente	Texto	Sólo para descartar duplicados
2	ESTABLECIMIENTO_SALUD	Código del establecimiento	Número	Incluido como posible predictor
3	GLOSA_ESTABLECIMIENTO_SALUD	Nombre del establecimiento	Texto	No incluido en modelo
4	PERTENENCIA_ESTABLECIMIENTO_SALUD	Tipo de pertenencia (perteneciente o no perteneciente al Sistema Nacional de Servicios de Salud [SNS])	Texto	Incluido como posible predictor
5	SEREMI	Código de la Secretaría Regional Ministerial de Salud (SEREMI)	Número	Incluido como posible predictor
6	SERVICIO_DE_SALUD	Código del servicio de salud	Número	Incluido como posible predictor
7	SEXO	Código del sexo biológico del paciente	Número	Incluido como posible predictor
8	FECHA_NACIMIENTO	Fecha de nacimiento del paciente	Fecha	No incluido en el modelo
9	EDAD_CANT	Registro numérico de la edad del paciente al ingreso	Número	Incluido como posible predictor
10	TIPO_EDAD	Unidad de medida de la edad, según modalidad descrita en valores	Número	No incluido en el modelo
11	EDAD_AÑOS	Edad en años del paciente al momento del ingreso	Número	No incluido en el modelo
12	PUEBLO_ORIGINARIO	Código del pueblo originario	Número	No incluido en el modelo
13	PAIS_ORIGEN	Código del país de origen	Número	No incluido en el modelo
14	GLOSA_PAIS_ORIGEN	Glosa del país de origen	Texto	Usado para excluir pacientes extranjeros
15	COMUNA_RESIDENCIA	Código de la comuna de residencia del paciente	Texto	Incluido como posible predictor
16	GLOSA_COMUNA_RESIDENCIA	Nombre de la comuna de residencia	Texto	No incluido en el modelo
17	REGION_RESIDENCIA	Código de la región de residencia del paciente	Texto	Incluido como posible predictor
18	GLOSA_REGION_RESIDENCIA	Nombre de la región de residencia	Texto	No incluido en el modelo
19	PREVISION	Código de la previsión de salud del paciente al momento del ingreso	Número	Incluido como posible predictor
20	BENEFICIARIO	Código de la clase beneficiario de Fonasa	Texto	Incluido como posible predictor
21	MODALIDAD	Código de la modalidad de atención de Fonasa	Número	Incluido como posible predictor
22	PROCEDENCIA	Código de la procedencia del paciente al momento del ingreso	Número	No incluido en el modelo
25	ANO_EGR	Año del egreso	Número	No incluido en el modelo
26	FECHA_EGR	Fecha de egreso	Fecha	No incluido en el modelo

Tabla 1 (Continued)

Ítem de la base de datos de egresos hospitalarios del DEIS				
N°	Nombre de la variable	Descripción	Tipo de dato	Utilizado en Modelo
27	AREA_FUNCIONAL_EGRESO	Código del nivel de cuidado o área funcional del que egresó el paciente	Número	Incluido como posible predictor
28	DIAS_ESTAD	Días de estadía total	Número	Variable objetivo
29	CONDICION_EGRESO	Código de la condición al egreso del paciente	Número	Usado para excluir egresos en condición de fallecidos
30	DIAG1	Código de la Clasificación Internacional de Enfermedades, 10.ª revisión (CIE-10), del diagnóstico principal	Texto	Incluido como posible predictor
31	GLOSA_DIAG1	Glosa del diagnóstico principal	Texto	Incluido como posible predictor
32	DIAG2	Código de la causa externa	Texto	No incluido en el modelo
33	GLOSA_DIAG2	Glosa de la causa externa	Texto	No incluido en el modelo
34	INTERV_Q	Código de la intervención quirúrgica	Número	Usado para excluir egresos sin cirugía asociada
35	CODIGO_INTERV_Q_PPAL	Código Fonasa de la intervención quirúrgica principal	Texto	Usado para identificar casos
36	GLOSA_INTERV_Q_PPAL	Glosa de la intervención quirúrgica principal	Texto	Incluido como posible predictor
37	PROCED	Código del procedimiento	Número	No incluido en el modelo
38	CODIGO_PROCED_PPAL	Código Fonasa del procedimiento principal	Texto	No incluido en el modelo
39	GLOSA_PROCED_PPAL	Glosa del procedimiento principal	Texto	No incluido en el modelo
*40	% POBREZA COMUNA	Porcentaje de pobreza de la comuna de origen	Número	Incluido como posible predictor

Nota: *Dato obtenido en la página web del Ministerio de Desarrollo Social.

Como se menciona, en la medida que la cantidad de ecuaciones o reglas de decisión generadas por los algoritmos aumenta, es típicamente esperable que el desempeño predictivo del algoritmo mejore. Sin embargo, el aumentar la complejidad del modelo sumando ecuaciones o reglas aumenta también la dificultad de interpretación humana de los modelos creados. Por lo anterior, también es posible agrupar a los algoritmos en “cajas abiertas” o “cajas cerradas”. De acuerdo a esta clasificación, los algoritmos de regresión logística, clasificador de árbol de decisión, máquina de vector de soporte, y clasificador bayesiano ingenuo son considerados más del tipo “cajas abiertas”, pues generan de menos a más ecuaciones según el orden en el que fueron listados, y los algoritmos bosques aleatorios, *adaboost*, y perceptrón multicapa, como “cajas cerradas”, pues generan de menos a más reglas de decisión según el orden en el que fueron listados.

Además, debido a su buen nivel de desempeño en otras tareas similares de clasificación binaria, se incluyó una

familia adicional de algoritmos denominada árboles de potenciación del gradiente, la cual también pertenecería al grupo de “cajas cerradas”, que genera un gran número de reglas computacionales, y que se implementó por medio del paquete XGBoost (una biblioteca de programas de código abierto).

El modelo se entrenó utilizando el 80% de los datos disponibles, y el 20% restante fue reservado para confirmar las capacidades predictivas del modelo. A esta parte de los datos se le denomina tradicionalmente muestra de prueba. De manera adicional, se realizó un proceso de remuestreo, o *bootstrapping*, de cien iteraciones, de manera a poder obtener intervalos de confianza de las cifras de ajuste y desempeño de los modelos seleccionados.

Evaluación y Ajuste de los Modelos

Para evaluar el desempeño de los algoritmos y modelos predictivos, utilizamos su poder de discriminación (cuantificada como el área bajo la curva de característica

operativa del receptor [*area under the receiver operating characteristic curve*, AUC_ROC,²⁵ en inglés]) en los datos.

La métrica de optimización se evaluó y ordenó en un *ranking* utilizando AUC-ROC, que corresponde a cuan bien un modelo puede distinguir entre dos grupos. El nivel de discriminación se catalogó como excelente (0.9–1), bueno (0.8–0.89), justo (0.7–0.79), pobre (0.6–0.69), y fallido (0.5–0.59).²⁶

También se reportan otras métricas tradicionales para problemas de clasificación: “exactitud”: razón del número correcto de predicciones sobre el total de muestras; “promedio de precisión”: promedio de exactitud de predicciones basado en el porcentaje de predicciones positivas que son correctas; “precisión”: medida de exactitud de una predicción basada en el porcentaje de predicciones positivas que sean correctas; “recall”: medidas del porcentaje de predicciones positivas contra posibles positivos en el conjunto de datos de datos; y “F1”: promedio armónico de precisión y recuerdo, con el mejor valor siendo 1 (precisión perfecta), y el peor, 0. Para cada una de las anteriores, además se reportan sus intervalos de confianza estimados con base en el procedimiento de remuestreo.

Reporte del Modelo

En este trabajo, el reporte del modelo utiliza las recomendaciones internacionales para este tipo de estudios,^{27,28} con la lista de verificación del informe transparente de un modelo de predicción multivariable para pronóstico o diagnóstico individual (*transparent reporting of a multivariable prediction model for individual prognosis or diagnosis*, TRIPOD, en inglés).²⁸

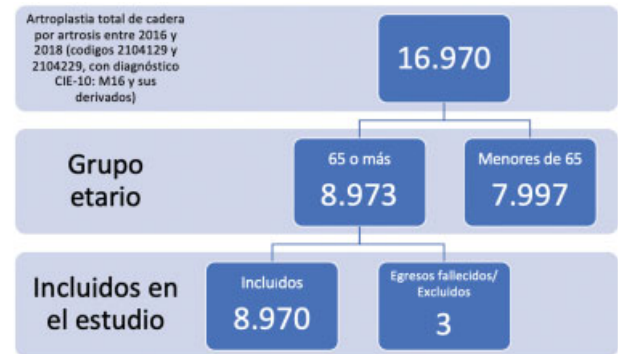


Fig. 1 Artroplastia total de cadera por artrosis entre 2016 y 2018 (códigos 2104129 y 2104229, con diagnóstico CIE-10: M16 y sus derivados).

Resultados

Se incluyeron finalmente 8.970 casos (► **Figura 1**): 5.662 mujeres (63,12%) y 3.308 (36,88%) hombres. La edad mediana era de 72 años, con un rango intercuartil de 9 años, y un rango entre 65 y 97 años (► **Figura 2**).

La muestra se compone por 6.746 (75,21%) pacientes de Fonasa, 1.599 (17,82%) pacientes de instituciones de salud previsual (Isapres), y 625 (6,97%) pacientes de otras previsiones. De los pacientes FONASA, 286 (4,2%) eran beneficiarios tipo A; 4.801 (71,2%), beneficiarios tipo B; 469 (6,9%), beneficiarios tipo C; y 1.191 (13,3%) beneficiarios tipo D. En este mismo grupo de pacientes FONASA, 5.321 (78,9%) se operaron bajo la modalidad de atención institucional, y 1.425 (21,1%), por la modalidad de libre elección.

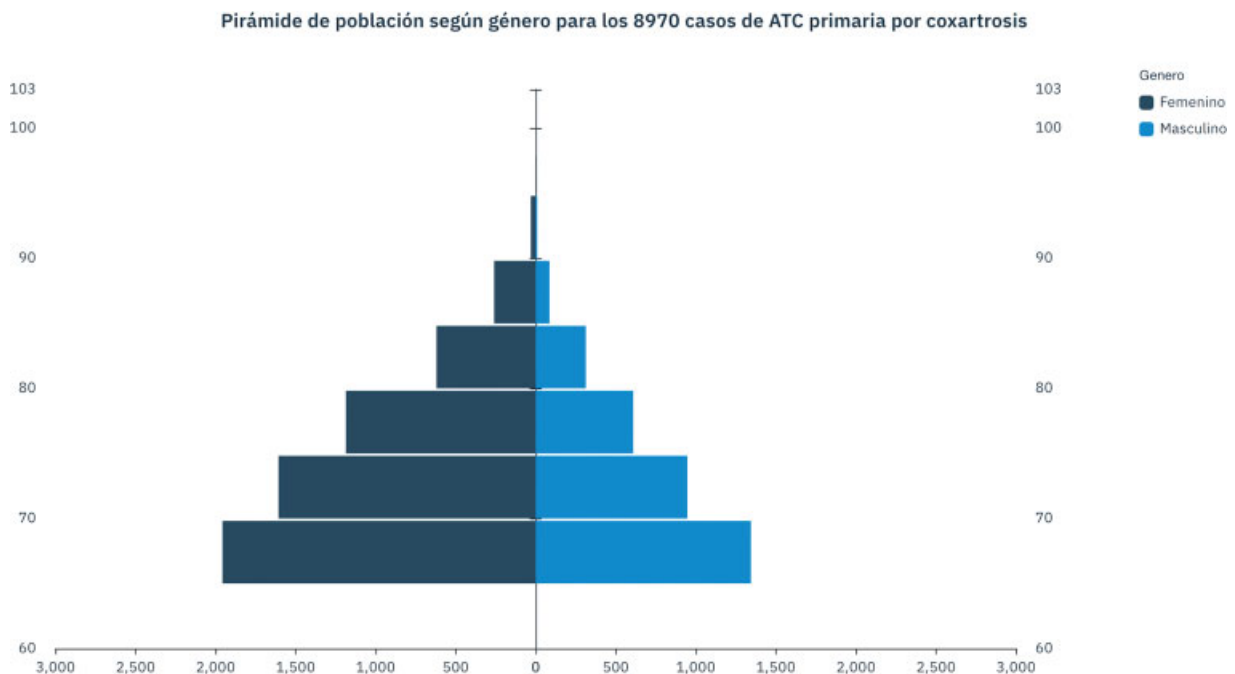


Fig. 2 Pirámide de población según género para los 8.970 casos de ATC primaria por coxartrosis.

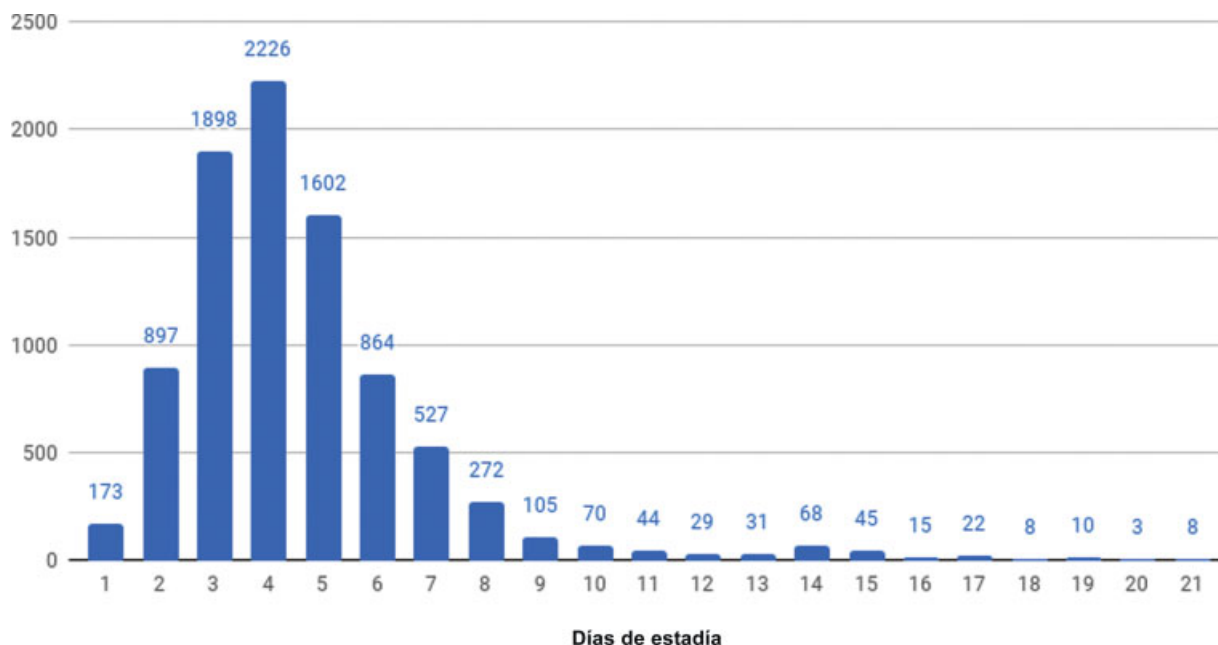


Fig. 3 Días de estadía.

Los 4 diagnósticos más frecuentes fueron M169 (6.124 casos; 68.27%), M161 (1.623 casos; 18.09%), M160 (862 casos; 9.61%), y M167 (176 casos; 1.96%).

Las 5 comunas de origen del paciente más frecuentes fueron Las Condes (426 casos; 4,75%), Viña del Mar (365 casos; 4,07%), La Florida (253 casos; 2,82%), Puente Alto (239 casos; 2,66%), y Santiago (235 casos; 2,62%), sumando entre estas 5 el 16.92% del total de casos en Chile.

Cien centros hospitalarios operaron pacientes de ATC por artrosis en el período estudiado. En total, 5.133 (81,88%) casos se operaron en centros pertenecientes al Sistema Nacional de Servicios de Salud, y 1.136 (18,12%) casos se operaron en centros privados.

La mediana de días de estadía fue 4, con un rango intercuartil de 2 días y un rango entre 1 y 143 días. El histograma de días de estadía se encuentra graficado en la ►Figura 3.

Los días de estadía categorizados por tipo de hospital y previsión se grafican en la ►Figura 4.

En total, 2.968 (33,09%) pacientes presentaron una estadía acortada, y 6.002 (66,91%) presentaron una estadía prolongada.

Desempeño de los Algoritmos de Decisión

Ocho algoritmos fueron evaluados con relación a los datos del entrenamiento y prueba; sin embargo, éstos fueron ordenados en un ranking de acuerdo con su desempeño en la muestra de prueba. Esta última se considera una mejor medida del desempeño del modelo al ser aplicado en escenarios reales. Entre ellos, el algoritmo XGBoost obtuvo el mejor desempeño, con una AUC-ROC promedio de 0,86 (DE: 0,0087). Esto quiere decir que el algoritmo XGBoost confirió el mejor desempeño al discriminar entre estadías hospitalarias corta y prolongada (mayor o menor de tres

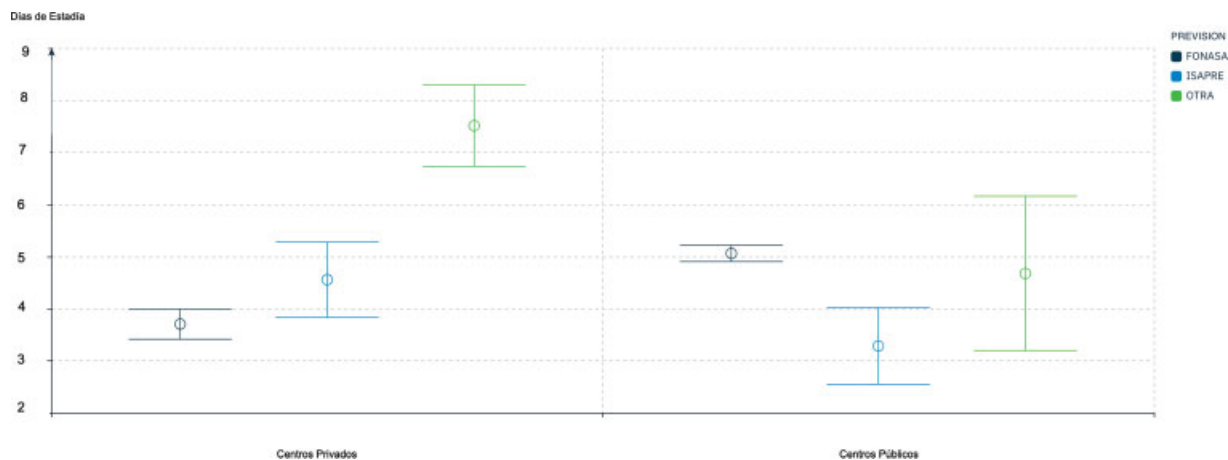


Fig. 4 Días de estadía por previsión y tipo de centro hospitalario.

Tabla 2 Resultados del ajuste de los modelos en la muestra de entrenamiento y prueba

Resultados en la muestra de entrenamiento										
Remuestreo de 100 muestras. Se reporta desviación estandar entre paréntesis										
	Exactitud total	Recall de clase 0	Recall de clase 1	Precisión de clase 0	Precisión de clase 1	Puntaje f1 0	Puntaje f1 1	Área bajo la curva		
XGBoost - árboles de potenciación de gradiente	81,56% (0,86%)	77,44% (1,40%)	86,05% (1,34%)	84,76% (1,20%)	79,24% (1,00%)	80,92% (0,94%)	82,50% (0,85%)	90,46% (0,77%)		
Máquina de vector de soporte	81,19% (0,38%)	78,76% (0,62%)	83,94% (0,68%)	83,07% (0,57%)	79,81% (0,44%)	80,86% (0,39%)	81,82% (0,39%)	89,55% (0,27%)		
AdaBoost	79,65% (0,43%)	76,79% (0,75%)	83,11% (0,93%)	81,98% (0,74%)	78,17% (0,47%)	79,30% (0,41%)	80,56% (0,45%)	88,16% (0,27%)		
Regresión logística	81,13% (0,42%)	78,32% (0,61%)	84,37% (0,79%)	83,37% (0,68%)	79,56% (0,44%)	80,76% (0,42%)	81,89% (0,45%)	89,62% (0,27%)		
Bosques aleatorios	79,40% (1,15%)	74,91% (2,07%)	83,68% (1,88%)	82,15% (1,62%)	76,96% (1,44%)	78,34% (1,37%)	80,16% (1,20%)	86,99% (0,91%)		
Red neural - perceptrón multicapa	89,99% (0,50%)	91,03% (1,21%)	88,79% (0,69%)	89,04% (0,57%)	90,84% (1,09%)	90,02% (0,62%)	89,80% (0,54%)	97,19% (0,31%)		
Árbol de decisión	66,04% (2,33%)	63,32% (27,95%)	68,33% (25,14%)	74,35% (14,06%)	70,46% (10,91%)	61,45% (13,47%)	64,69% (8,31%)	74,05% (2,03%)		
Bayesiano ingenuo	65,07% (1,60%)	38,05% (3,89%)	94,97% (0,68%)	88,33% (0,89%)	60,56% (1,38%)	53,07% (3,81%)	73,94% (0,89%)	67,51% (1,73%)		
Resultados en la muestra de prueba										
Remuestreo de 100 muestras. Se reporta desviación estandar entre paréntesis										
	Exactitud total	Recall de clase 0	Recall de clase 1	Precisión de clase 0	Precisión de clase 1	Puntaje f1 0	Puntaje f1 1	Área bajo la curva		
XGBoost - árboles de potenciación de gradiente	81,74% (0,87%)	75,62% (1,60%)	80,23% (2,24%)	88,56% (1,19%)	61,97% (1,73%)	81,56% (0,92%)	69,90% (1,31%)	86,01% (0,87%)		
Máquina de vector de soporte	81,35% (0,37%)	77,21% (1,40%)	78,81% (1,98%)	88,05% (1,08%)	63,12% (1,86%)	82,26% (0,90%)	70,07% (1,48%)	85,68% (0,86%)		
AdaBoost	79,95% (0,40%)	75,81% (1,33%)	79,98% (1,81%)	88,45% (0,99%)	62,06% (1,61%)	81,63% (0,83%)	69,87% (1,26%)	85,55% (0,90%)		
Regresión logística	81,34% (1,60%)	76,60% (3,89%)	78,49% (0,68%)	87,81% (0,89%)	62,40% (1,38%)	81,81% (3,81%)	69,51% (0,89%)	85,16% (1,73%)		

Tabla 2 (Continued)

	(0,43%)	(1,33%)	(1,88%)	(1,03%)	(1,73%)	(0,87%)	(1,39%)	(0,90%)
Bosques aleatorios	79,30%	72,70%	77,43%	86,70%	58,43%	79,06%	66,56%	82,32%
	(1,23%)	(2,32%)	(2,88%)	(1,54%)	(2,33%)	(1,56%)	(2,04%)	(1,36%)
Red neural - perceptrón multicapa	89,91%	82,12%	64,44%	82,37%	64,07%	82,24%	64,23%	82,07%
	(0,58%)	(1,16%)	(2,43%)	(1,13%)	(1,77%)	(0,81%)	(1,70%)	(0,95%)
Árbol de decisión	65,82%	62,70%	66,65%	83,63%	53,84%	66,05%	54,06%	72,58%
	(2,47%)	(28,09%)	(25,86%)	(8,78%)	(12,33%)	(17,75%)	(4,52%)	(2,15%)
Bayesiano ingenuo	66,51%	36,80%	90,04%	88,14%	41,39%	51,81%	56,69%	64,35%
	(1,70%)	(4,05%)	(1,36%)	(1,63%)	(1,73%)	(4,14%)	(1,59%)	(1,94%)

días). En segundo lugar, observamos que el algoritmo lineal de máquina de vector de soporte obtuvo una AUC-ROC muy cercana de 0,8568 (DE: 0,0086), pero con una DE menor.

En la ►Tabla 2, se muestran las diferentes métricas de clasificación para cada uno de los algoritmos evaluados. Siguiendo el concepto de exactitud (razón del número correcto de predicciones sobre el total de muestras), el algoritmo XGBoost fue capaz de predecir correctamente el 81,74% de las veces cuando un caso correspondía a una estadía corta o prolongada.

Para indagar además respecto de la importancia relativa de las variables explicativas, se reporta en la ►Figura 5 el puntaje de importancia asignado por el algoritmo a las treinta variables más importantes. En este sentido, destaca el hecho de que la región de residencia, el servicio de salud, el establecimiento de salud donde se operó el paciente, y la modalidad de atención son las variables que más determinan el tiempo de estadía de un paciente.

La figura 6 muestra un árbol de clasificación representativo del algoritmo XGBoost.

Discusión

En el presente trabajo, se desarrolló y validó exitosamente un modelo de predicción de estadía hospitalaria en pacientes chilenos mayores de 65 años sometidos a ATC utilizando inteligencia artificial en su modalidad de aprendizaje de máquinas y macrodatos de origen nacional. El algoritmo XGBoost confirmó el mejor desempeño predictivo al discriminar cuando la estadía hospitalaria se clasifica en estadía acortada y prolongada (mayor o menor de tres días). Además, se encontró que los cinco factores más importantes en esta predicción, todos de acceso libre en la base de datos ministerial, son la *región de residencia, el servicio de salud, el establecimiento de salud, y la modalidad*. La exactitud de clasificación del algoritmo es buena.

Según Ramkumar et al.,²⁹ el aprendizaje de máquinas se podría describir como un programa que desarrolla tareas automáticamente desde una fuente de datos sin una programación explícita. Esta tecnología rápidamente se ha incorporado a la medicina, y representa la extensión natural de los métodos estadísticos tradicionales. Específicamente en la literatura de artroplastia, hay varios trabajos recientes que utilizan aprendizaje de máquinas para crear modelos de predicción de estadía hospitalaria y pagos relacionados a las cirugías,²⁹ de probabilidad de complicaciones,²⁶ de satisfacción³⁰ etc. Todos estos trabajos, al igual que el presente, utilizan amplias bases de datos que pueden ser consideradas macrodatos.³¹

Nuestro trabajo tiene varias limitaciones y algunos aspectos destacables. La primera limitación es que se trata de un trabajo de registro; por lo tanto, existe la posibilidad de problemas de recolección y codificación que finalmente pudieran alterar los resultados, especialmente considerando que se usan los códigos CIE-10 y Fonasa para identificar los casos estudiados. A pesar de ello, creemos que, por tratarse de una base de datos ministerial, con toda la

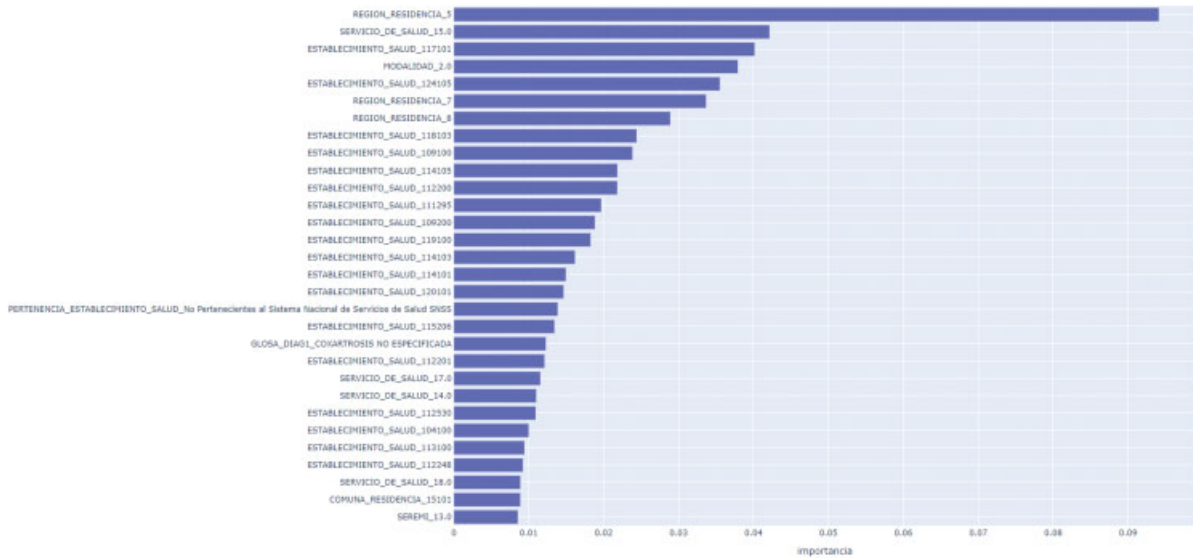


Fig. 5 Importancia relativa de las 30 variables más importantes del modelo explicativo de estadia hospitalaria.

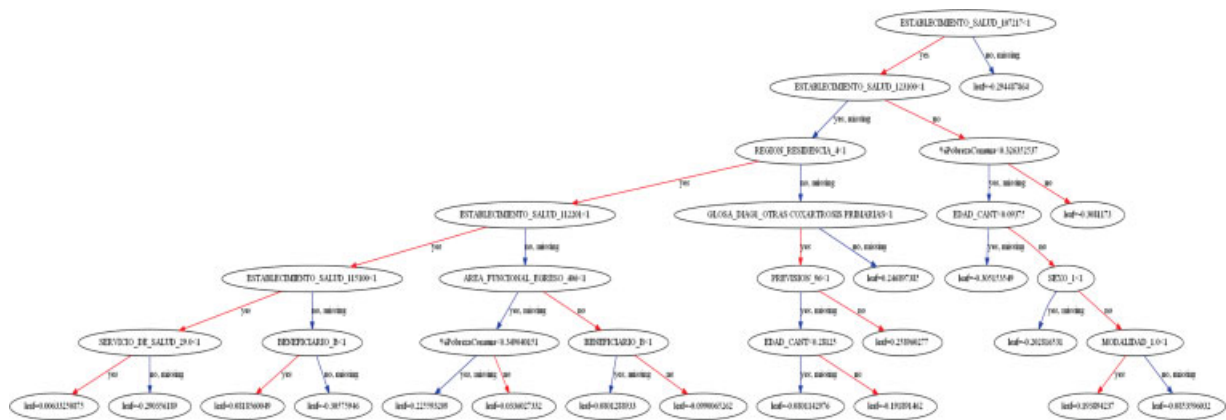


Fig. 6 Arbol de clasificaci'on representativo del algoritmo XGBoost.

rigurosidad que ello implica, es lo suficientemente sólida como para sobreponerse a esta limitación. Segundo, ninguno de los trabajos de base de datos captura suficiente información al nivel de pacientes.³² Esto es especialmente importante en nuestro trabajo, considerando que la mayoría de los trabajos realizados en el hemisferio norte que utilizan esta metodología emplean variables a nivel del paciente, incluyendo comorbilidades y, en algunos casos, funcionalidad.^{16,26,30} Consideramos que esta es la principal falencia de nuestro trabajo; sin embargo, la base de datos utilizada es la única que nos permita acceder libremente a macrodatos a nivel nacional. Pese a esta observación, es necesario destacar que el rol de las características individuales del paciente pudiera no ser el más relevante en explicar la estadia hospitalaria en artroplastia electiva. Kang et al.³³ demostraron, en una serie de dos mil pacientes, que los principales determinantes de estadia prolongada en artroplastia son sociales: la admisión al hospital el día previo a la cirugía, y el empezar tardíamente con la rehabilitación postoperatoria. Paralelamente, Burn et al.³⁴ demostraron

que, si bien los factores individuales de los pacientes son relevantes para explicar la estadia hospitalaria en artroplastia, entre 1997 y 2014 en el Reino Unido, se lograron menores estadias hospitalarias por la mejora en la eficiencia de las practicas, dado que el perfil de los pacientes operados permaneció estable. Reforzando más aún el hecho de que las características individuales de los pacientes son secundarias al explicar la variabilidad en el momento del alta hospitalaria, el Cleveland Clinic OME Arthroplasty Group demostró (utilizando macrodatos estadounidenses), en pacientes sometidos a ATC electiva que “mientras los factores relacionados al paciente explican alguna variación en la estadia hospitalaria, los principales responsables son los factores relacionados al procedimiento, específicamente el hospital”³⁵ donde se operó el paciente, teniendo además un rol protagónico el abordaje quirúrgico empleado. Esta evidencia mencionada ayuda a entender los resultados de nuestro trabajo y a sopesar la falta de variables individuales como limitación no crítica de nuestro modelo. Tercero, teniendo en cuenta que la pandemia COVID-19 pudiera

haber influenciado la práctica de las ATCs¹¹ en Chile en cuanto a su postoperatorio y su alta hospitalaria más precoz,^{12,36} creemos que los datos correspondientes a los años 2016-2018 pueden no ser completamente representativas del escenario que vamos a vivir el año 2021. Sin embargo, los fundamentos de nuestro algoritmo pueden utilizarse para evaluar los resultados de altas hospitalarias tras ATC registrados para el año 2020 y los venideros.

La pregunta que se alza es: ¿tiene este calculador utilidad en nuestro medio? La evaluación de posibilidad de alta precoz o tardía de una cirugía altamente frecuente y garantizada por ley es de total relevancia en políticas públicas. El calcular las diferentes posibilidades de alta precoz que tiene un paciente Fonasa que se opera en el hospital A versus el hospital B, o clínica X, es útil para visualizar la variabilidad que existe en las prácticas. Al momento de generar modelos tipo pagos “paquetizados”, es importante predecir si el paciente operado en el hospital A va a tener una hospitalización más prolongada que en el hospital B. La utilidad del calculador “al lado de la cama” puede ser limitada por la ausencia de macrodatos clínicos de libre acceso en Chile, pero, por otra parte, la utilidad con perspectiva de evaluación de desempeño de instituciones es altísima. Como expusimos en los objetivos del trabajo, la identificación de grupos con alta probabilidad de estadia acortada (ciertos pacientes en algunos hospitales) puede ayudar a éstas a mejorar aún más sus prácticas. En el otro extremo, el identificar centros hospitalarios que, contando con los mismos recursos, no son eficientes en sus estadias hospitalarias, puede ayudarlos a mejorar.

Entre las fortalezas de nuestro trabajo, creemos que la primera y más importante es el logro de un trabajo multidisciplinario entre cuatro expertos, dos de ellos cirujanos y dos ingenieros con educación formal en inteligencia artificial, que lograron por primera vez realizar un trabajo de macrodatos e inteligencia artificial en nuestra especialidad en Chile.

Conclusión

En el presente estudio, se desarrollaron algoritmos de aprendizaje de máquinas basados en macrodatos chilenos de libre acceso, y se logró validar una herramienta que demuestra una adecuada capacidad discriminatoria para predecir la probabilidad de estadia hospitalaria acortada versus prolongada en adultos mayores sometidos a ATC por artrosis.

Referencias

- Problema de salud GES N° 12: Endoprótesis total de cadera en personas de 65 años y más con artrosis de cadera con limitación funcional severa. Orientación en Salud Superintendencia de Salud, Gobierno de Chile n.d. <http://www.supersalud.gob.cl/difusion/665/w3-article-586.html> (accessed December 18, 2020).
- Grosso MJ, Neuwirth AL, Boddapati V, Shah RP, Cooper HJ, Geller JA. Decreasing Length of Hospital Stay and Postoperative Complications After Primary Total Hip Arthroplasty: A Decade Analysis From 2006 to 2016. *J Arthroplasty* 2019;34(03): 422–425. Doi: 10.1016/j.arth.2018.11.005
- Goyal N, Chen AF, Padgett SE, et al. Otto Aufranc Award: A Multicenter, Randomized Study of Outpatient versus Inpatient Total Hip Arthroplasty. *Clin Orthop Relat Res* 2017;475(02): 364–372. Doi: 10.1007/s11999-016-4915-z
- Paredes O, Ñuñez R, Klaber I. Successful initial experience with a novel outpatient total hip arthroplasty program in a public health system in Chile. *Int Orthop* 2018;42(08):1783–1787. Doi: 10.1007/s00264-018-3870-6
- Greenky MR, Wang W, Ponzio DY, Courtney PM. Total Hip Arthroplasty and the Medicare Inpatient-Only List: An Analysis of Complications in Medicare-Aged Patients Undergoing Outpatient Surgery. *J Arthroplasty* 2019;34(06):1250–1254. Doi: 10.1016/j.arth.2019.02.031
- Featherall J, Brigati DP, Faour M, Messner W, Higuera CA. Implementation of a Total Hip Arthroplasty Care Pathway at a High-Volume Health System: Effect on Length of Stay, Discharge Disposition, and 90-Day Complications. *J Arthroplasty* 2018;33(06):1675–1680. Doi: 10.1016/j.arth.2018.01.038
- Ripollés-Melchor J, Abad-Motos A, Díez-Remesal Y, et al; Postoperative Outcomes Within Enhanced Recovery After Surgery Protocol in Elective Total Hip and Knee Arthroplasty (POWER2) Study Investigators Group for the Spanish Perioperative Audit and Research Network (REDGERM) Association Between Use of Enhanced Recovery After Surgery Protocol and Postoperative Complications in Total Hip and Knee Arthroplasty in the Postoperative Outcomes Within Enhanced Recovery After Surgery Protocol in Elective Total Hip and Knee Arthroplasty Study (POWER2). *JAMA Surg* 2020;155(04): e196024–e196024. Doi: 10.1001/jamasurg.2019.6024
- Manning DW, Edelstein AI, Alvi HM. Risk Prediction Tools for Hip and Knee Arthroplasty. *J Am Acad Orthop Surg* 2016;24(01): 19–27. Doi: 10.5435/JAAOS-D-15-00072
- Sconza C, Respizzi S, Grappiolo G, Monticone M. The Risk Assessment and Prediction Tool (RAPT) after Hip and Knee Replacement: A Systematic Review. *Joints* 2019;7(02):41–45. Doi: 10.1055/s-0039-1693459
- Díaz Ledezma C, Radovic I. What's new in hip arthroplasty? South American perspective. *Recent Advances in Orthopedics-2*. Jaypee Brothers Medical Publishers (P) Ltd; 2018
- Parvizi J, Gehrke T, Krueger CA, et al; International Consensus Group (ICM) and Research Committee of the American Association of Hip and Knee Surgeons (AAHKS) Resuming Elective Orthopaedic Surgery During the COVID-19 Pandemic: Guidelines Developed by the International Consensus Group (ICM). *J Bone Joint Surg Am* 2020;102(14):1205–1212. Doi: 10.2106/JBJS.20.00844
- Donell ST, Thaler M, Budhiparama NC, et al. Preparation for the next COVID-19 wave: The European Hip Society and European Knee Associates recommendations. *Knee Surg Sports Traumatol Arthrosc* 2020;28(09):2747–2755. Doi: 10.1007/s00167-020-06213-z
- Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial Intelligence and Orthopaedics: An Introduction for Clinicians. *J Bone Joint Surg Am* 2020;102(09):830–840. Doi: 10.2106/JBJS.19.01128
- Cabitz F, Locoro A, Banfi G. Machine Learning in Orthopedics: A Literature Review. *Front Bioeng Biotechnol* 2018;6:75. Doi: 10.3389/fbioe.2018.00075
- Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* 2018;33(08): 2358–2361. Doi: 10.1016/j.arth.2018.02.067
- Haerberle HS, Helm JM, Navarro SM, et al. Artificial Intelligence and Machine Learning in Lower Extremity Arthroplasty: A Review. *J*

- Arthroplasty 2019;34(10):2201–2203. Doi: 10.1016/j.arth.2019.05.055
- 17 Ramkumar PN, Haeberle HS, Bloomfield MR, et al. Artificial Intelligence and Arthroplasty at a Single Institution: Real-World Applications of Machine Learning to Big Data, Value-Based Care, Mobile Health, and Remote Patient Monitoring. *J Arthroplasty* 2019;34(10):2204–2209. Doi: 10.1016/j.arth.2019.06.018
 - 18 Anderson AB, Grazal CF, Balazs GC, Potter BK, Dickens JF, Forsberg JA. Can Predictive Modeling Tools Identify Patients at High Risk of Prolonged Opioid Use After ACL Reconstruction? *Clin Orthop Relat Res* 2020;478(07):0–1618. Doi: 10.1097/CORR.0000000000001251
 - 19 Departamento de Estadísticas e Información de Salud n.d. <https://deis.minsal.cl/#publicaciones> (accessed December 18, 2020).
 - 20 Farley KX, Anastasio AT, Premkumar A, Boden SD, Gottschalk MB, Bradbury TL. The Influence of Modifiable, Postoperative Patient Variables on the Length of Stay After Total Hip Arthroplasty. *J Arthroplasty* 2019;34(05):901–906. Doi: 10.1016/j.arth.2018.12.041
 - 21 Observatorio Social - Ministerio de Desarrollo Social y Familia. n.d. <http://observatorio.ministeriodesarrollosocial.gob.cl/pobrezacomunal-2017#basedatos> (accessed February 4, 2021).
 - 22 Mackinnon A. The use and reporting of multiple imputation in medical research - a review. *J Intern Med* 2010;268(06):586–593. Doi: 10.1111/j.1365-2796.2010.02274.x
 - 23 López V, Fernández A, García S, Palade V, Herrera F. An Insight into Classification with Imbalanced Data: Empirical Results and Current Trends on Using data Intrinsic Characteristics. *Inf Sci* 2013;250:113–141. Doi: 10.1016/j.ins.2013.07.007
 - 24 del Rio S, Benitez JM, Herrera F. Analysis of Data Preprocessing Increasing the Oversampling Ratio for Extremely Imbalanced Big Data Classification. 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom, vol. 2, 2015, p. 180–5. <https://doi.org/10.1109/Trustcom.2015.579>
 - 25 Cerda J, Cifuentes L. Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. *Rev Chilena Infectol* 2012;29(02):138–141. Doi: 10.4067/S0716-10182012000200003
 - 26 Harris AHS, Kuo AC, Weng Y, Trickey AW, Bowe T, Giori NJ. Can Machine Learning Methods Produce Accurate and Easy-to-use Prediction Models of 30-day Complications and Mortality After Knee or Hip Arthroplasty? *Clin Orthop Relat Res* 2019;477(02):452–460. Doi: 10.1097/CORR.0000000000000601
 - 27 Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet* 2019;393(10181):1577–1579. Doi: 10.1016/S0140-6736(19)30037-6
 - 28 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594. Doi: 10.1136/bmj.g7594
 - 29 Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and Validation of a Machine Learning Algorithm After Primary Total Hip Arthroplasty: Applications to Length of Stay and Payment Models. *J Arthroplasty* 2019;34(04):632–637. Doi: 10.1016/j.arth.2018.12.030
 - 30 Kunze KN, Polce EM, Sadauskas AJ, Levine BR. Development of Machine Learning Algorithms to Predict Patient Dissatisfaction After Primary Total Knee Arthroplasty. *J Arthroplasty* 2020;35(11):3117–3122. Doi: 10.1016/j.arth.2020.05.061
 - 31 MeSH Browser. n.d. <https://meshb.nlm.nih.gov/record/ui?ui=D000077558> (accessed December 19, 2020).
 - 32 Grauer JN, Leopold SS. Editorial: large database studies—what they can do, what they cannot do, and which ones we will publish. *Clin Orthop Relat Res* 2015;473(05):1537–1539. Doi: 10.1007/s11999-015-4223-z
 - 33 Kang HW, Bryce L, Cassidy R, Hill JC, Diamond O, Beverland D. Prolonged length of stay (PLOS) in a high-volume arthroplasty unit. *Bone Jt Open* 2020;1(08):488–493. Doi: 10.1302/2633-1462.18.BJO-2020-0047.R1
 - 34 Burn E, Edwards CJ, Murray DW, et al. Trends and determinants of length of stay and hospital reimbursement following knee and hip replacement: evidence from linked primary care and NHS hospital records from 1997 to 2014. *BMJ Open* 2018;8(01):e019146. Doi: 10.1136/bmjopen-2017-019146
 - 35 Girbino KL, Klika AK, Barsoum WK, et al; Cleveland Clinic OME Arthroplasty Group. Understanding the Main Predictors of Length of Stay After Total Hip Arthroplasty: Patient-Related or Procedure-Related Risk Factors? *J Arthroplasty* 2021;36(05):1663–1670.e4. Doi: 10.1016/j.arth.2020.11.029
 - 36 Athey AG, Cao L, Okazaki K, et al. Survey of AAHKS International Members on the Impact of COVID-19 on Hip and Knee Arthroplasty Practices. *J Arthroplasty* 2020;35(7S):S89–S94. Doi: 10.1016/j.arth.2020.04.053