

Uso ético y sostenible de los datos clínicos del mundo real para la generación de nuevo conocimiento médico

XLV Jornadas de Economía de la Salud, Sevilla, 17 de junio, 2026

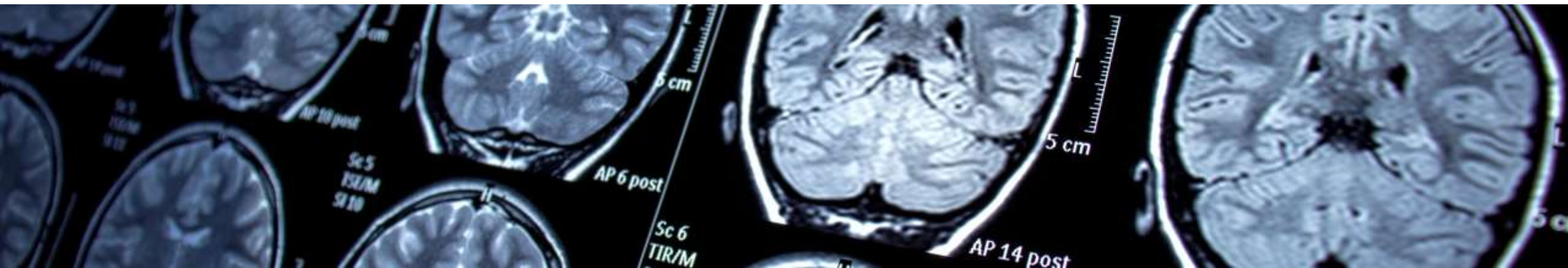
Joaquín Dopazo

Barcelonaβeta Brain Research Center (BBRC), Barcelona, Spain.

Hospital del Mar Research Institut Barcelona (HMRIB), Barcelona, Spain.

CIBER Enfermedades Raras, Co Director Programa de genómica y bioinformática

Académico correspondiente de la Real Academia Nacional de Medicina



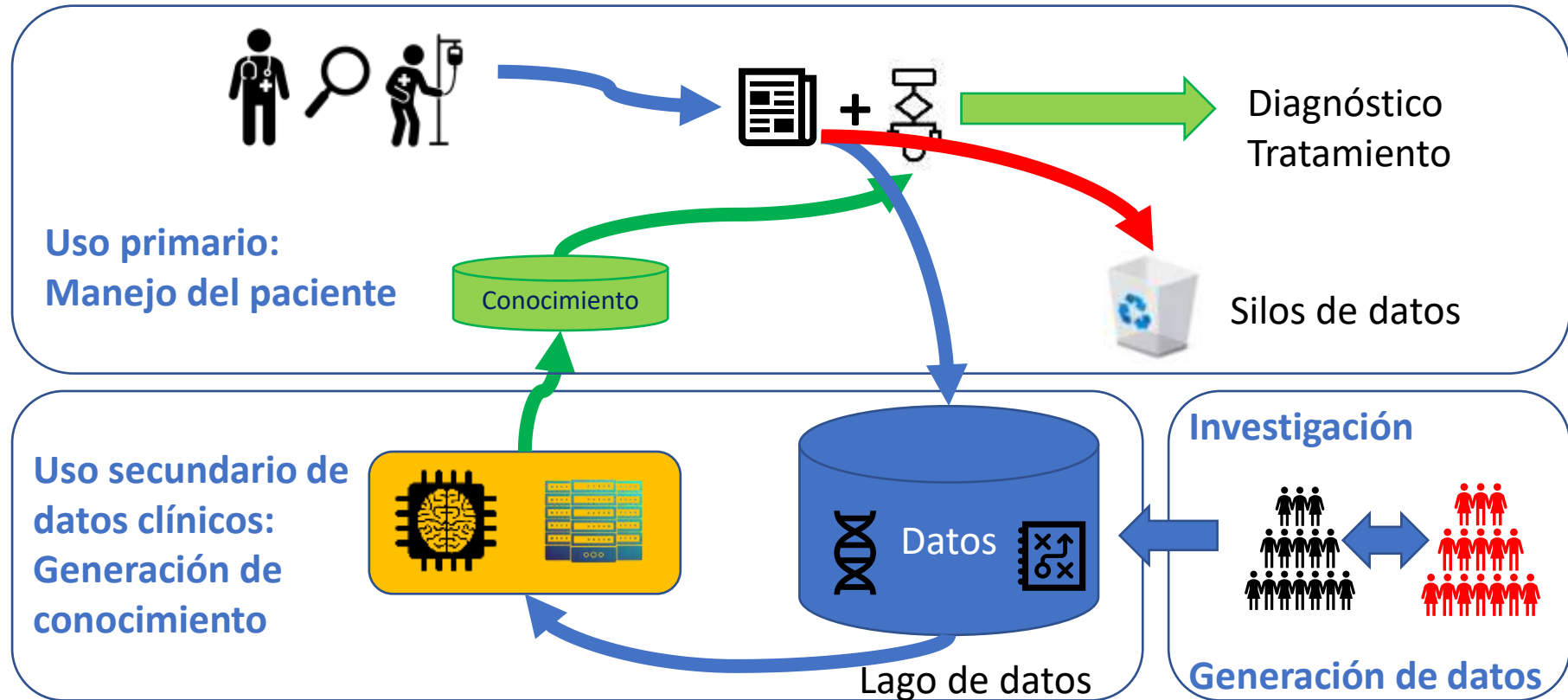
El ciclo de los datos



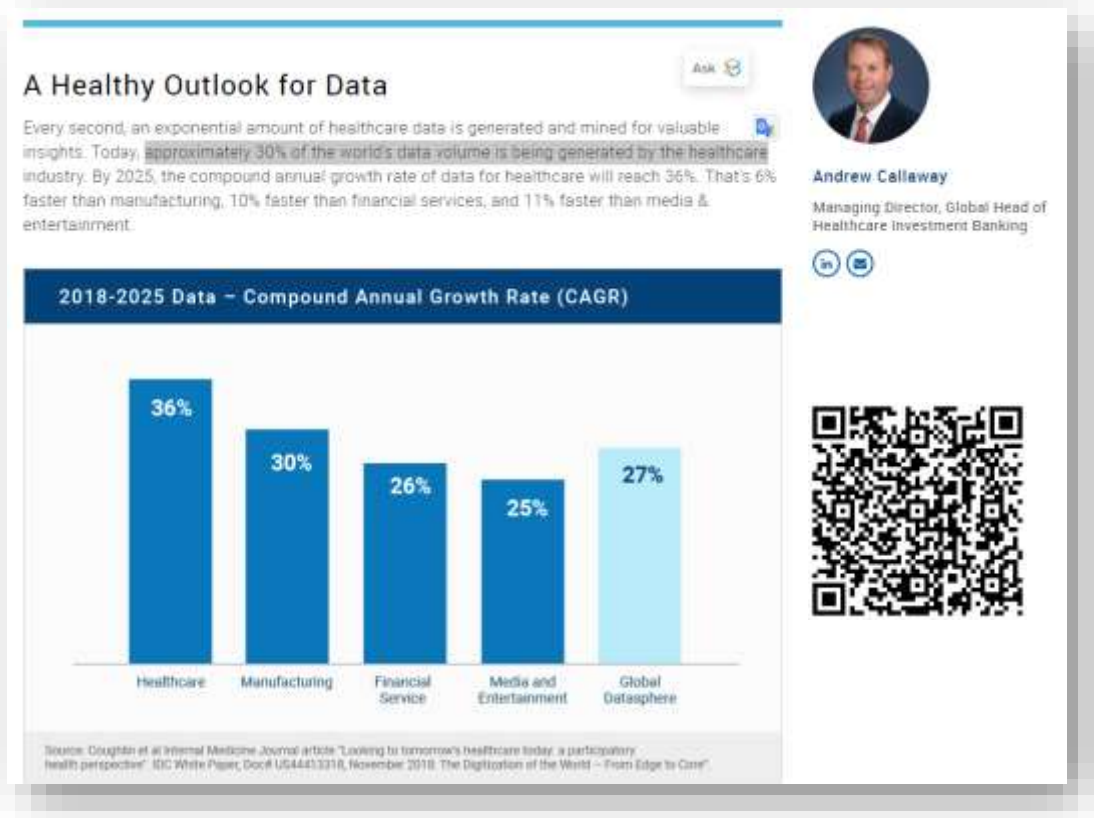
Artículo 5 de la RGPD: los datos se recogen con fines específicos, explícitos y legítimos (uso primario) y no serán tratados posteriormente de manera incompatible con dichos fines.

Excepciones de usos secundarios:

- Interés vital
- Salud pública
- Gestión
- **Investigación**



La magnitud de los datos clínicos



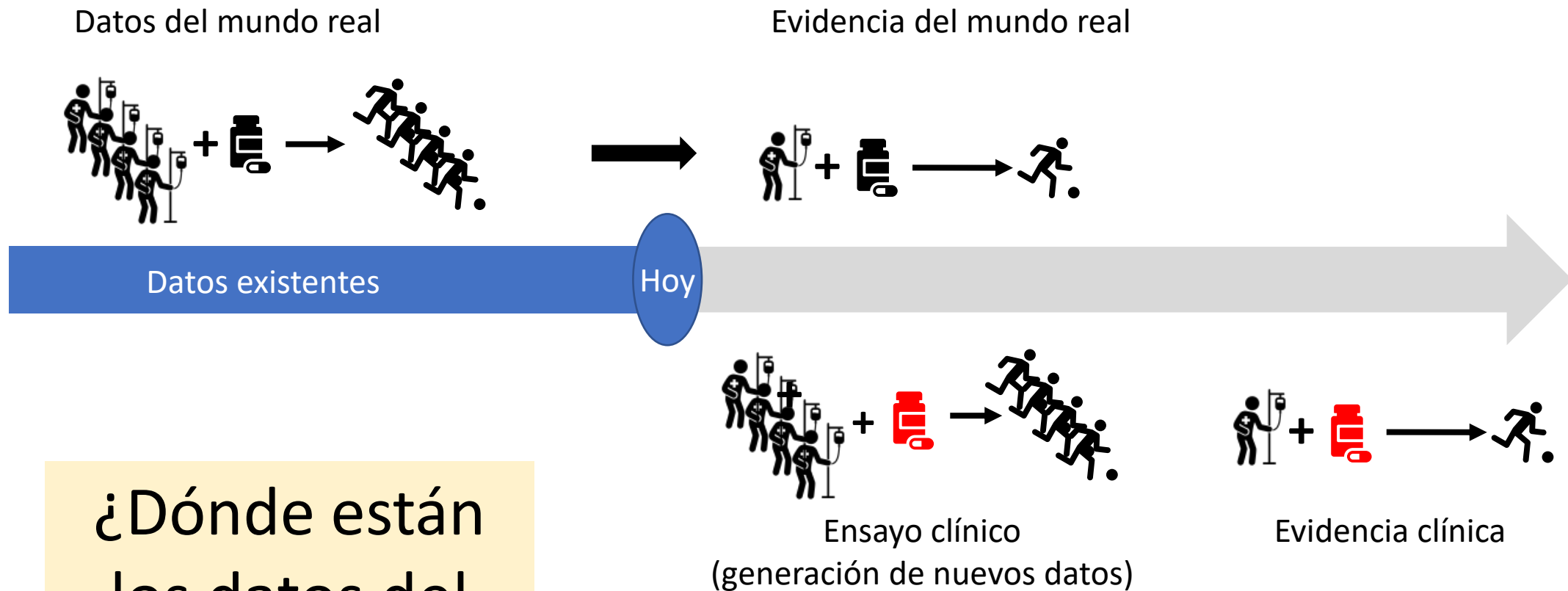
Aproximadamente 1/3 de los datos que existen en el mundo se han generado en un entorno hospitalario.

Son Datos del Mundo Real (DMR / RWD). En medicina, los datos derivados de diversas fuentes se asocian con resultados en una población heterogénea de pacientes en entornos reales, incluyendo, entre otros, historiales clínicos electrónicos y encuestas de pacientes. Los datos del mundo real se refieren a datos observacionales, a diferencia de los datos recopilados en un entorno experimental, como un ensayo clínico aleatorizado.

(Wikipedia)



Evidencia del mundo real y ensayos clínicos



¿Dónde están los datos del mundo real?

DMR:

- Datos recogidos por el hospital
- Tamaños muestrales enormes

EECC:

- Datos seleccionados para el estudio
- Tamaños muestrales pequeños

Lagos de datos regionales

Cambio de escala de estudios multicentro tradicionales (pocos hospitales) a estudios regionales (TODO el sistema hospitalario de la región).

Regiones del tamaño de países medianos europeos. Poblaciones de millones de habitantes, datos en el orden de miles de millones (diagnósticos, intervenciones, análisis, hospitalizaciones, urgencias, UCI, atención primaria y especializada, farmacia, vacunación, evaluación funcional, datos socioeconómicos, estilo de vida, constantes, etc.)



Base Poblacional de Salud (Andalucía)



AQUAS (Cataluña)



The Valencia Health System Integrated Database (Valencia)



Población: 8.6M, 16.5M de individuos acumulados



Población: 8.2M.



Población: 5.3M, 7Mde individuos acumulados

Consideraciones sobre el escenario convencional usado en la generación de conocimiento clínico



El sistema de salud genera datos y los custodia en un entorno seguro

Los datos se extraen del entorno seguro del sistema de salud y se analizan fuera

El sistema de salud queda relegado al papel de mero productor pasivo de datos y hasta debe de pagar por los productos desarrollados por terceros con sus datos.

Riesgos y desventajas del escenario convencional



La anonimización total no existe. Implica una degradación de datos. Debe ser hecha específicamente para cada estudio, de forma que se preserve al máximo la información sin comprometer identidades de pacientes



Reidentificación y pérdida de datos son riesgos reales.

A finales de 2016, se descifraron las identidades de los médicos en un conjunto de datos abiertos de registros de facturación médica australianos. Recientemente se reportó la aparición de datos de UK biobank en repositorios.



El mal uso de los datos y las listas negra existen:

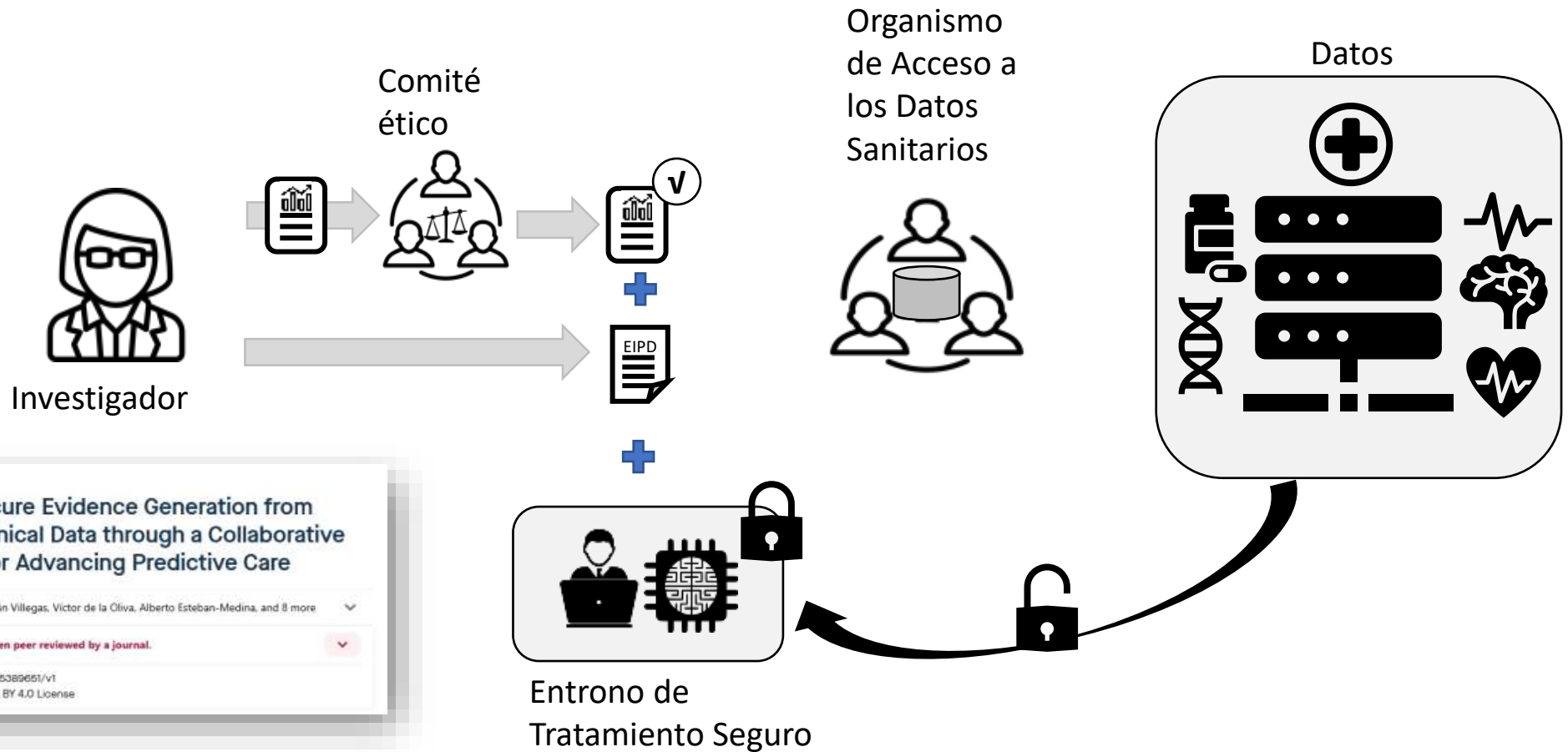
Aunque un acceso abierto de los investigadores a los datos acelera la generación de conocimiento, el riesgo de su mal uso es una realidad innegable y perjudicial para la población en general.



Concepto de justicia:

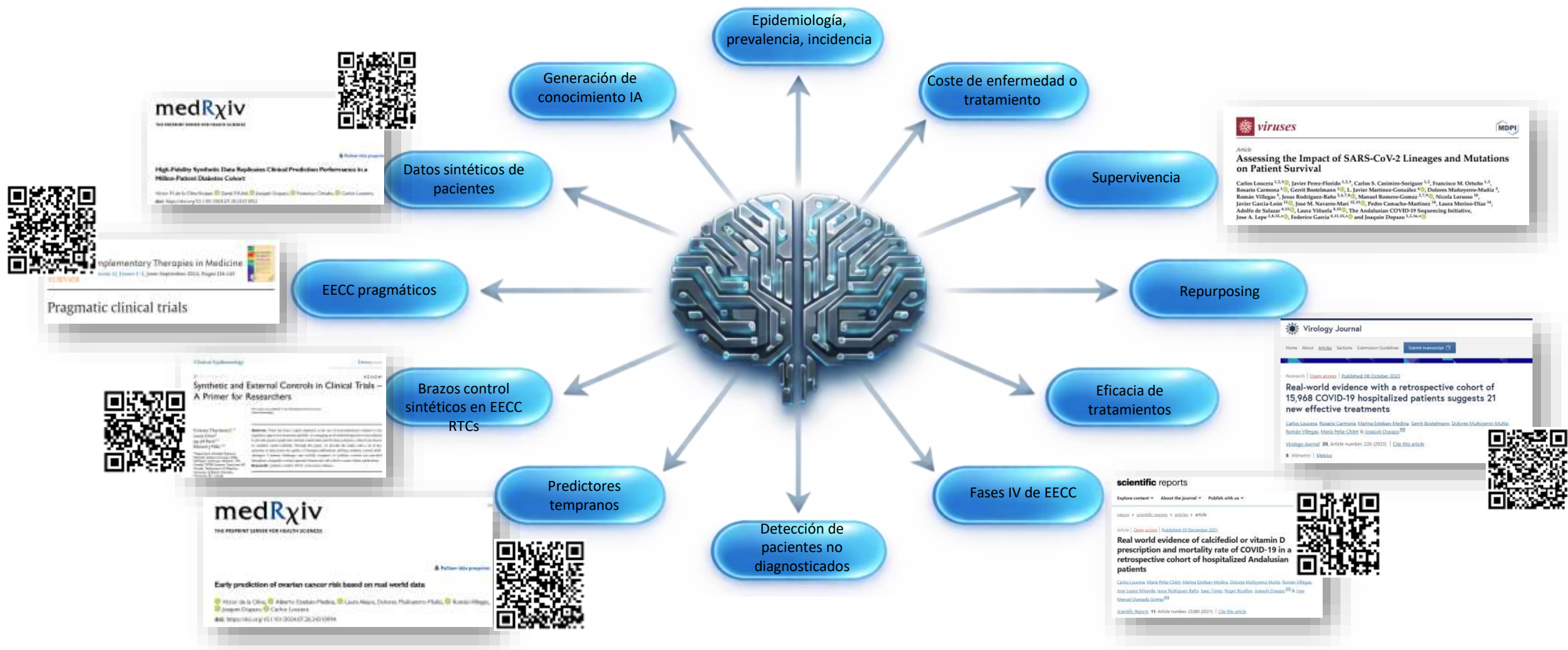
Los datos son del paciente y si generan riqueza debería de retornar a este. El paciente paga con sus impuestos un sistema de salud que genera datos suyos y los custodia (dos procesos caros).

Procedimiento para el uso de los datos (EEDS)



IESE-Novartis Award Initiatives to improve patient care

Los RWD permiten generar evidencia que los EECC no pueden cubrir solos



Algunos resultados obtenidos

viruses MDPI

Article

Assessing the Impact of SARS-CoV-2 Lineages and Mutations on Patient Survival

Carlos Loucera^{1,2,†}, Javier Pérez-Florido^{1,2,†}, Carlos S. Casimiro-Soriguer^{1,2}, Francisco M. Ortuño^{1,3}, Rosario Carmona^{1,3}, Gerit Bostelmann^{1,3}, L. Javier Martínez-González^{4,5}, Dolores Muñoz-Muñiz¹, Román Villegas⁶, Jesús Rodríguez-Baño^{2,6,7,8,9}, Manuel Romero-Gómez^{2,7,8,9}, Nicola Lorusso¹⁰, Javier García-León¹¹, Jose M. Navarro-Mari^{12,13}, Pedro Camacho-Martínez¹⁴, Laura Merino-Díaz¹⁴, Adolfo de Salazar^{4,14}, Laura Viñuela^{4,14}, The Andalusian COVID-19 Sequencing Initiative, Jose A. Lape^{2,4,14}, Federico García^{4,14,15} and Joaquín Dopazo^{1,2,14,*}

scientific reports

Explore content About the journal Publish with us

nature > scientific reports > articles > article

Article | Open access | Published: 01 December 2021

Real world evidence of calcifediol or vitamin D prescription and mortality rate of COVID-19 in a retrospective cohort of hospitalized Andalusian patients

Carlos Loucera, María Peña-Ched, Marina Esteban-Medina, Dolores Muñoz-Muñiz, Román Villegas, Jose Lopez-Miranda, Jesus Rodriguez-Baño, Jose Luis Lopez-Gatell, Manuel Romero-Gomez, Joaquin Dopazo & Jose Manuel Gonzalez Gomez

Scientific Reports 11, Article number: 23660 (2021) | Cite this article

Journal of Epidemiology and Global Health (2024) 14:1504–1517
https://doi.org/10.1007/s44197-024-00298-2

RESEARCH ARTICLE

Assessing COVID-19 Vaccine Effectiveness and Risk Factors for Severe Outcomes through Machine Learning Techniques: A Real-World Data Study in Andalusia, Spain

Álvaro Serrano-Ortiz^{1,2,3}, Juan Luis Romero-Cabrera^{4,5}, Jaime Monserrat Villatoro^{6,7}, Jaime Cordero-Ramos^{8,9,10}, Rafael Ruiz-Montero^{1,2,11}, Álvaro Ritoré¹², Joaquín Dopazo^{13,10}, Jorge del Diego Salas¹⁴, Valle García Sánchez^{15,16}, Inmaculada Salcedo-Lea^{1,2,11}, Migu

scientific reports

Explore content About the journal Publish with us

nature > scientific reports > articles > article

Article | Open access | Published: 29 May 2024

Aripiprazole as protector against COVID-19 mortality

C. Loucera-Muñiz, M. Canal-Rivero, M. Ruiz-Veguilla, B. Carmona, G. Bostelmann, N. Garrido-Torres, L. Dopazo & B. Crespo-Barceno

Scientific Reports 14, Article number: 12362 (2024) | Cite this article

Research Article

Ethical and Secure Evidence Generation from Regionwide Clinical Data through a Collaborative Environment for Advancing Predictive Care

Dolores Muñoz-Muñiz, Román Villegas, Victor de la Oliva, Alberto Esteban-Medina, and 8 more

This is a preprint; it has not been peer reviewed by a journal.

https://doi.org/10.21203/rs.3.rs-5389661/v1
This work is licensed under a CC BY 4.0 License

medRxiv THE PREPRINT SERVER FOR HEALTH SCIENCES

Early prediction of ovarian cancer risk based on real world data

Victor de la Oliva, Alberto Esteban-Medina, Laura Añejo, Dolores Muñoz-Muñiz, Román Villegas, Joaquín Dopazo, Carlos Loucera
doi: https://doi.org/10.1101/2024.07.24.24103994

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Real world data analysis of venous thromboembolism, adverse major cardiovascular events, neoplasia and serious infections in ulcerative colitis patients

Victor de la Oliva, Alberto Esteban-Medina, Patricia Fernández del Valle, Ana Sánchez, M. Belén Susin, Joaquín Dopazo, Carlos Loucera, Eduardo Leo-Camero
doi: https://doi.org/10.1101/2025.01.21.25320903

Virology Journal

Home About Articles Sections Submission Guidelines Submit manuscript

Research | Open access | Published: 06 October 2023

Real-world evidence with a retrospective cohort of 15,968 COVID-19 hospitalized patients suggests 21 new effective treatments

Carlos Loucera, Rosario Carmona, Marina Esteban-Medina, Gerit Bostelmann, Dolores Muñoz-Muñiz, Román Villegas, María Peña-Ched & Joaquín Dopazo

Virology Journal 20, Article number: 226 (2023) | Cite this article

BMJ Open Original research

Menopause age and hypercholesterolemia comorbidities: a region-wide retrospective cohort study in Andalusia, Spain (2016–2022)

Alberto Esteban-Medina,¹ Victor de la Oliva,¹ Patricia Fernández del Valle,¹ Ana Sánchez,¹ M Belen Susin,¹ Carlos Loucera,^{1,2} Guillermo Antiñolo,^{2,3,4,5} Joaquín Dopazo^{1,2,3}

BMJ Open Original research

Menopause age and hypercholesterolemia comorbidities: a region-wide retrospective cohort study in Andalusia, Spain (2016–2022)

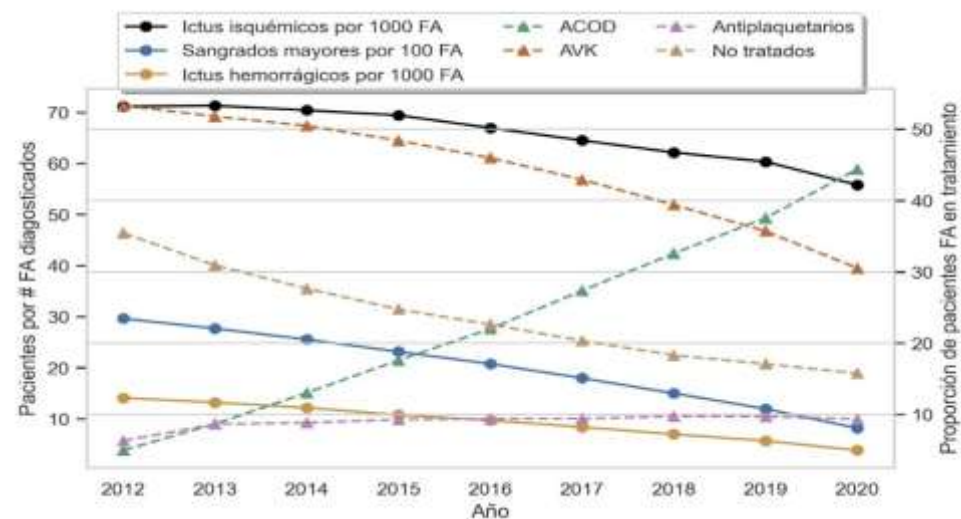
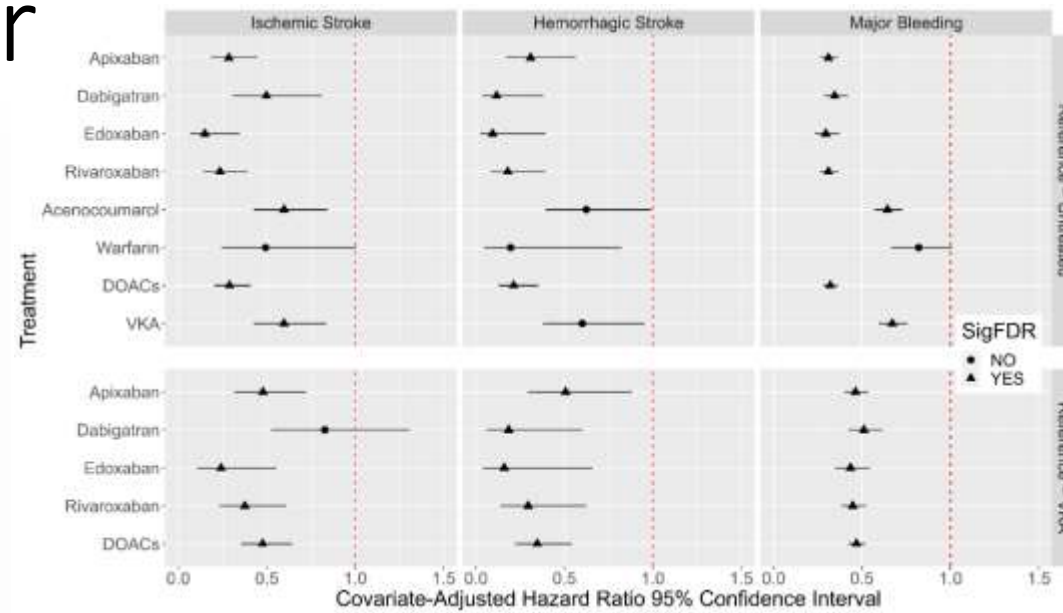
Alberto Esteban-Medina,¹ Victor de la Oliva,¹ Patricia Fernández del Valle,¹ Ana Sánchez,¹ M Belen Susin,¹ Carlos Loucera,^{1,2} Guillermo Antiñolo,^{2,3,4,5} Joaquín Dopazo^{1,2,3}

Caso de uso: Riesgo de ictus en pacientes con fibrilación auricular

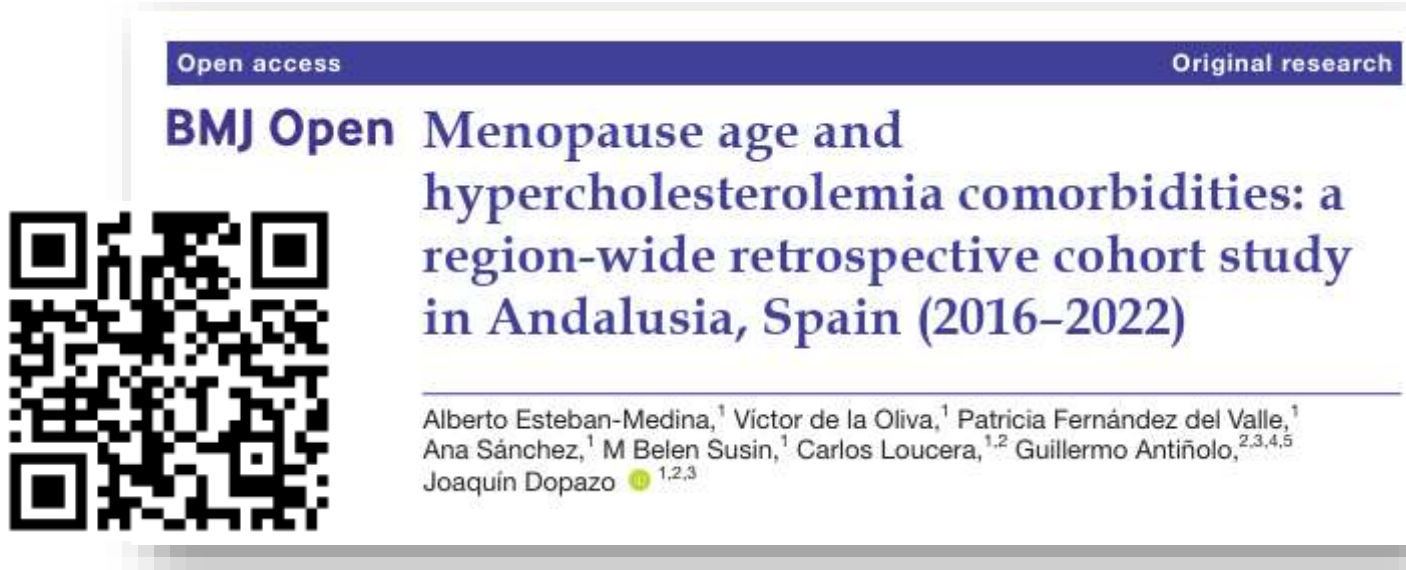
Original article
 Evidence of the association between increased use of direct oral anticoagulants and a reduction in the rate of atrial fibrillation-related stroke and major bleeding at the population level (2012–2019)
 Carlos Loucera^{a,b}, Rosario Carmona^{a,b,c}, Gerrit Bostelmann^a, Dolores Muñozerro-Muñiz^d, Román Villegas^d, Rafael Gonzalez-Manzanares^e, Joaquin Dopazo^{a,b,c,f,1}, Manuel Anguita^{e,x,1}

^a Computational Medicine Platform, Andalusian Public Foundation Progress and Health-FPS, Sevilla, Spain
^b Institute of Biomedicine of Seville (IBIS), Hospital Virgen del Rocío, Sevilla, Spain
^c Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), FPS, Hospital Virgen del Rocío, Sevilla, Spain
^d Subdirección Técnica Asesora de Gestión de la Información, Servicio Andaluz de Salud, Sevilla, Spain
^e Department of Cardiology, Hospital Universitario Reina Sofía, Instituto Maimónides para la Investigación Biomédica de Córdoba, Universidad de Córdoba, Córdoba, Spain
^f FPS/ELIXIR-ES, Fundación Progreso y Salud (FPS), CDCA, Hospital Virgen del Rocío, Sevilla, Spain

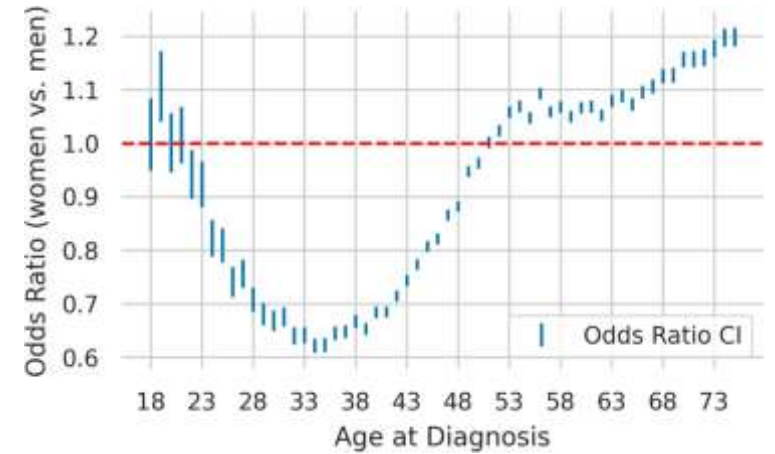
Pacientes mayores de 18 años con diagnóstico de **Fibrilación Auricular** antes de Enero de 2021. En ellos se recogió la **historia de enfermedades**, mediante la clasificación CIE-10, así como los **tratamientos** antitrombóticos (ACO y antiagregantes) de interés y las fechas de prescripción de dichos fármacos. Un total de **64.538** pacientes cumplieron los criterios de inclusión y pasaron los controles de calidad.



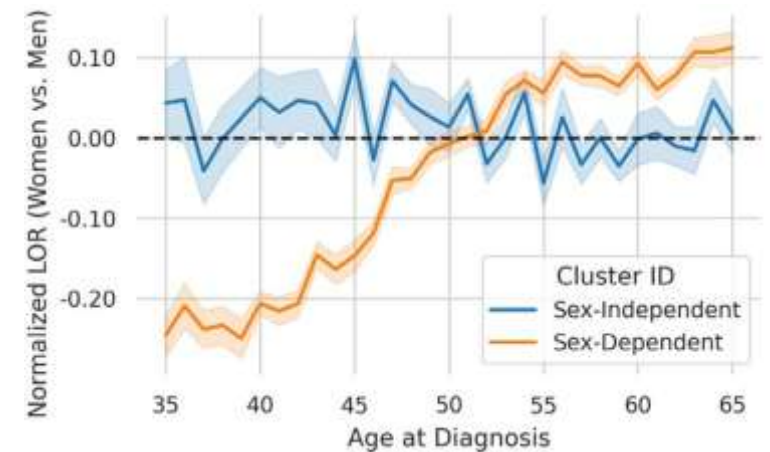
El impacto de la menopausia en la hipercolesterolemia y otras comorbilidades: un estudio poblacional



Se analizó una cohorte retrospectiva de **557 034** pacientes diagnosticados con hipercolesterolemia entre 2016 y 2022. Los pacientes se estratificaron por sexo y grupos de edad (<50 y ≥50 años). Se realizó un análisis comparativo para evaluar las diferencias en la prevalencia de la hipercolesterolemia y las comorbilidades asociadas.

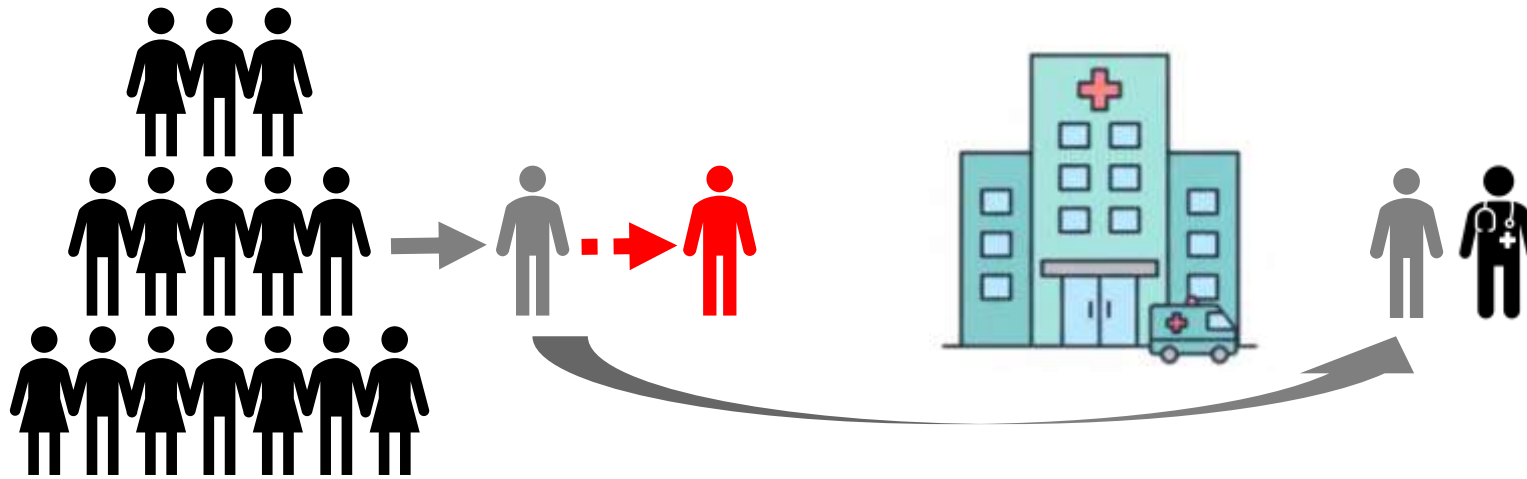


Odd ratios e intervalos de confianza (IC) de la incidencia total de comorbilidades en mujeres frente a hombres.



Log-OR ajustados de la incidencia de comorbilidades en mujeres frente a hombres. Se representan dos grupos de enfermedades.

Más vale prevenir que curar



Prevención primaria: en población general, antes de que aparezca la enfermedad (**cribado**)

Prevención secundaria: en población de riesgo o afectada, antes de que aparezca el síntoma



Ventajas para al paciente

- Mejor pronóstico y supervivencia.
- Intervenciones menos invasivas.
- Mayor calidad de vida.



Ventajas para el sistema de salud

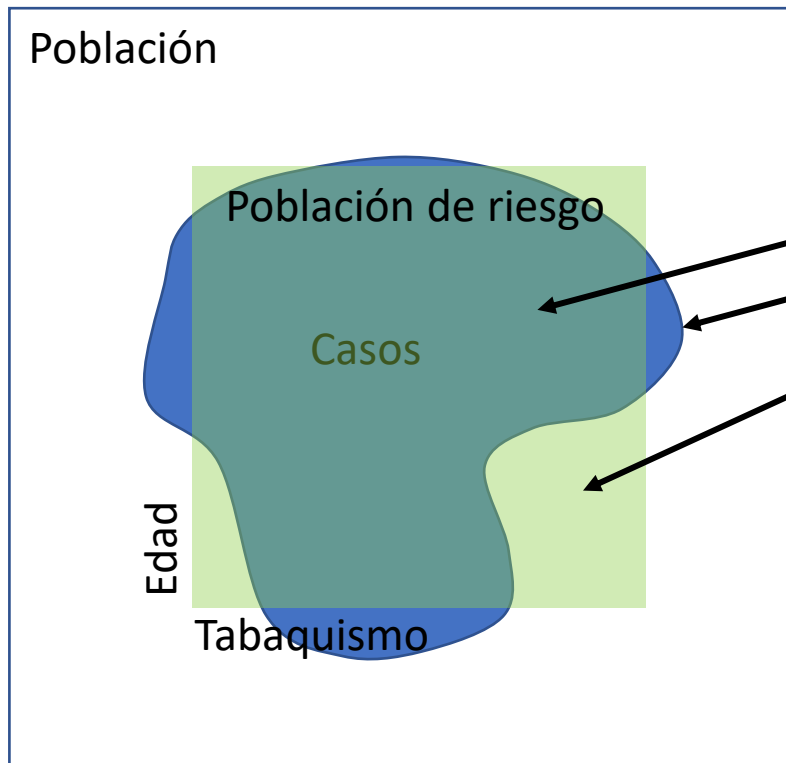
- Menos costes evitables.
- Menos hospitalizaciones.
- Sostenibilidad a largo plazo.



Ventajas para la planificación

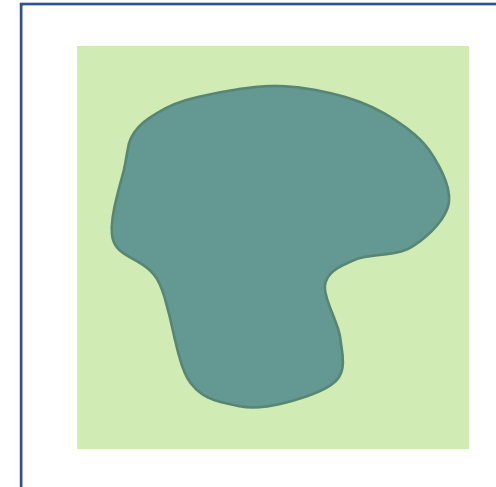
- Anticipación a la demanda.
- Priorización de recursos.
- Decisiones basadas en riesgo.

Definición convencional de la población de riesgo, especificidad y exactitud



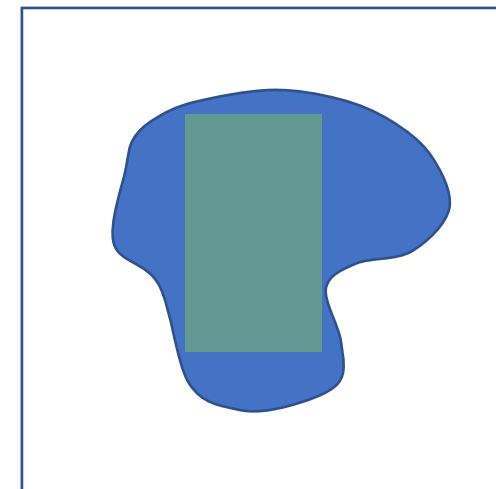
Acertos
Falsos negativos
Falsos positivos

El coste y efectividad de un cribado depende de la relación entre los aciertos, los falsos positivos y los falsos negativos (teniendo en cuenta la adherencia)



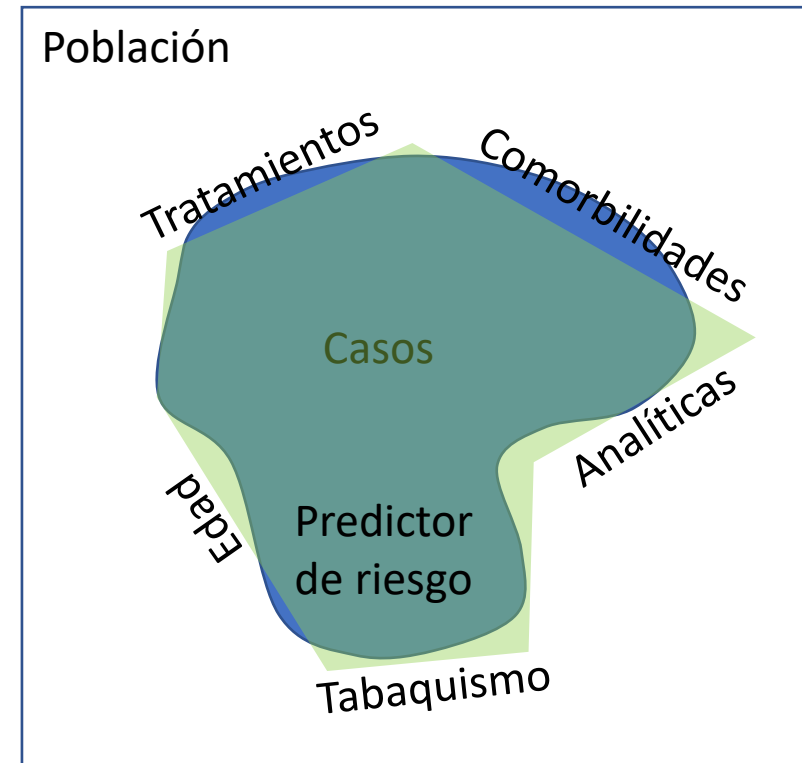
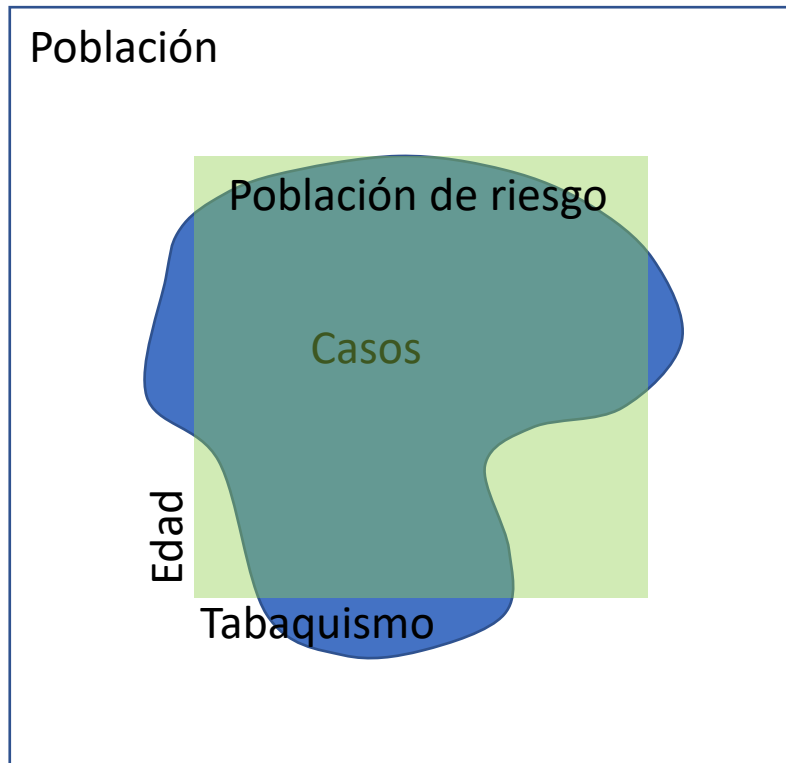
Coste / efectividad

Cribado: muy **preciso** (diagnostica a todos) pero poco específico (muchos diagnosticados no son de riesgo)

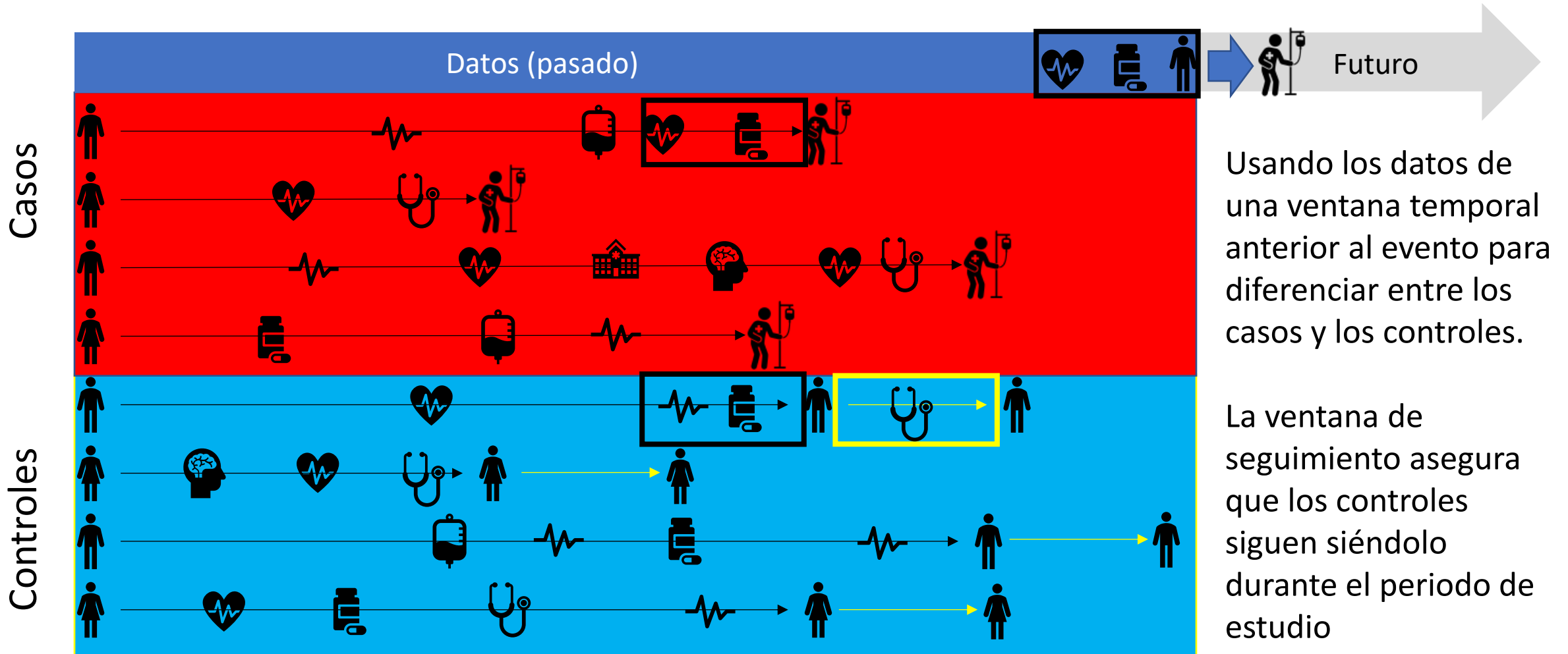


Score de riesgo muy **específico** (siempre acierta) pero poco preciso (se deja muchos por diagnosticar)

Prevención de precisión: ¿Podemos ser más precisos que un cribado etario?



¿Como se hace un predictor temprano de riesgo?



No predecimos con datos del diagnóstico, predecimos con la huella clínica previa al diagnóstico

Predictor temprano de cáncer de ovario



 CONSEJERÍA DE SALUD Y FAMILIAS
Secretaría General de Investigación, Desarrollo e Innovación en Salud

DICTAMEN DEL COMITÉ COORDINADOR DE ÉTICA DE LA INVESTIGACIÓN BIOMÉDICA DE ANDALUCÍA

D. Joaquín Alanís López, Secretario del Comité Coordinador de Ética de la Investigación Biomédica de Andalucía

CERTIFICA

Que este Comité, en virtud de lo previsto en la Ley 14/2007, de 3 de julio, de Investigación Biomédica y el Decreto 8/2020, de 30 de enero, por el que se regulan los órganos de ética asistencial y de la investigación biomédica en Andalucía, ha evaluado la propuesta del promotor: FUNDACIÓN PROGRESO Y SALUD el proyecto de investigación titulado: "ESTUDIO OBSERVACIONAL RETROSPECTIVO PARA EL DESARROLLO DE PREDICTORES TEMPRANOS DE CÁNCER DE OVARIO", Protocolo versión 3 de fecha 4 de abril de 2022, con código de solicitud: DOCAI y con código interno (FEIBA) de protocolo: 0298-N-22, cuyo investigación principal es: D. Joaquín Dopazo Blázquez.

Y CONSIDERA QUE:

Se cumplen los requisitos necesarios de idoneidad del protocolo en relación con los objetivos del estudio y se ajusta a los principios éticos aplicables a este tipo de estudios y recogidos en la Declaración de Helsinki.

El tratamiento de los datos de carácter personal de los participantes se ajusta a lo dispuesto en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 y en la Ley Orgánica 3/2018, de 5 de diciembre de Protección de Datos Personales y garantía de los derechos digitales y de igual forma se ajusta a las normas de buena práctica clínica.

Están justificados los riesgos y molestias previsibles para los participantes. Es adecuado el procedimiento para obtener el consentimiento informado.

Por todo lo anterior:

El Comité Coordinador de Ética de la Investigación Biomédica de Andalucía, en su reunión del día 29 de marzo de 2022 (Acta 3/22) tras la evaluación del citado estudio emite un **DICTAMEN FAVORABLE**.

Dicho Comité, está constituido y actúa de acuerdo con la normativa vigente y las directrices de la Conferencia Internacional de Buena Práctica Clínica.



Código Seguro de Verificación: UZ2PUL53PFC0JUS6C942N-EZCL. Permite la verificación de la integridad de una copia de este documento electrónico en la dirección: <https://verifed.sistemasinformatica.juntadeandalucia.es/verifed/>

FIRMACIO POR:	Joaquín Alanís López	FECHA:	07/04/2022
ID FEIBA:	UZ2PUL53PFC0JUS6C942N-EZCL	PÁGINA:	1/2

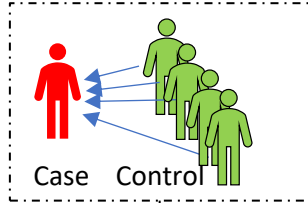


- Cáncer de ovario con incidencia de 1 en 78 mujeres. Síntomas inespecíficos. Alta mortalidad (75%) al descubrimiento, normalmente tardío. Detección temprana con tasa de curación del 90%)
- Predictor temprano basado únicamente en los datos que deja una mujer por su paso por el sistema de salud (visitas a primaria o especializada, analíticas, urgencias, etc.)
- Cohorte retrospectiva de 7.353 mujeres con diagnóstico de cáncer de ovario

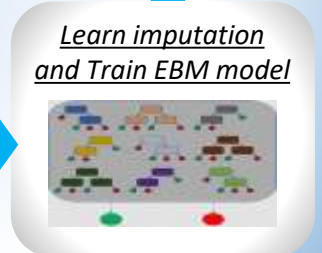
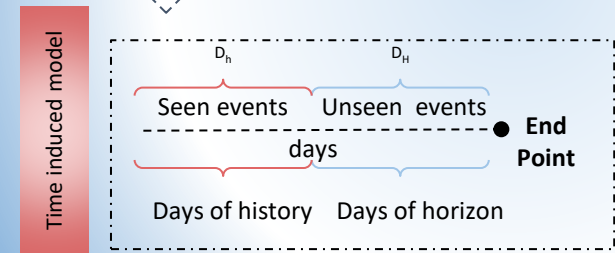
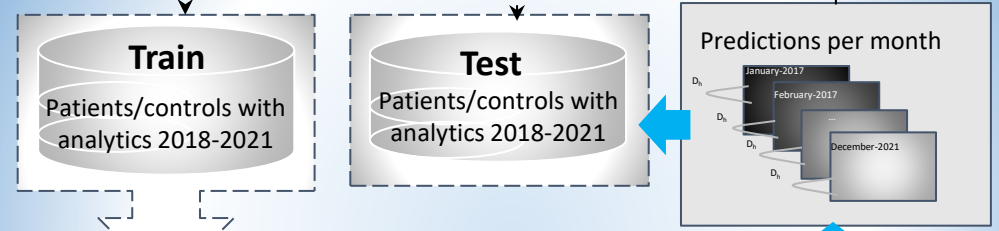
El *Comité Coordinador de Ética de la Investigación Biomédica de Andalucía*, en su reunión del día 29 de marzo de 2022 (Acta 3/22) tras la evaluación del citado estudio emite un **DICTAMEN FAVORABLE**

Modelado y validación

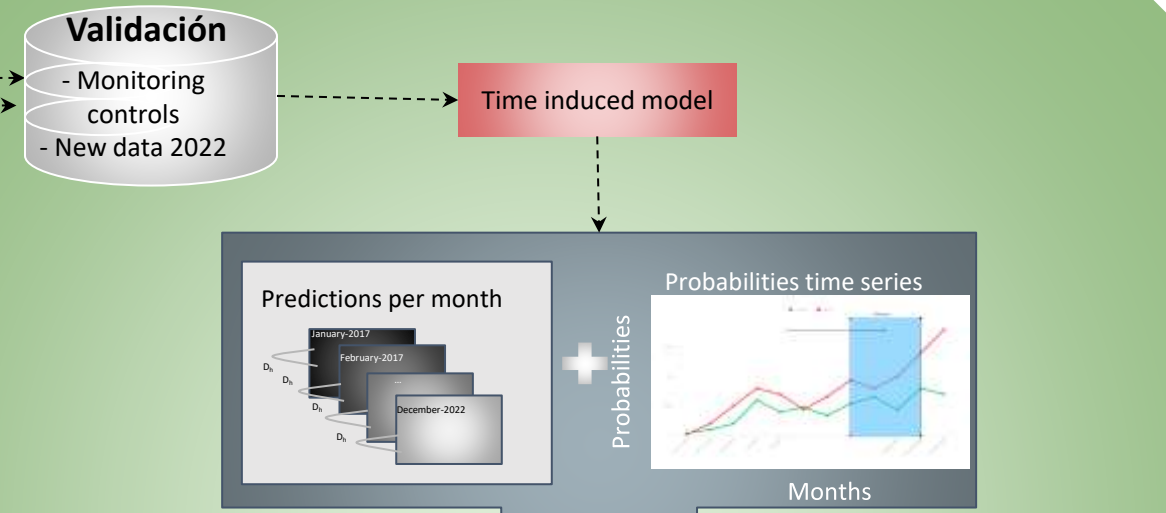
- I. Enfermedades crónicas
- II. Historial diagnósticos
- III. Consultas Especialistas
- IV. Analíticas
- V. Etc.



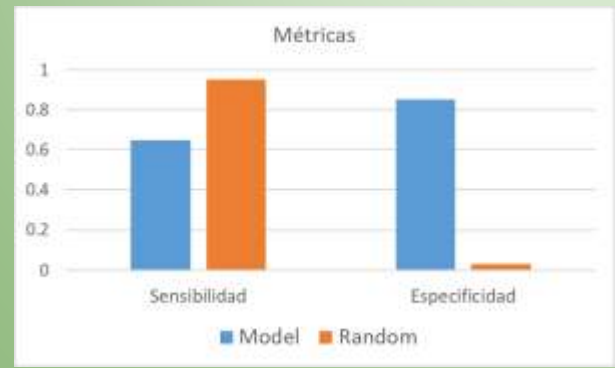
(2018-2021)



Validación



Results



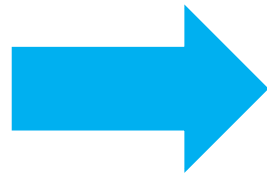
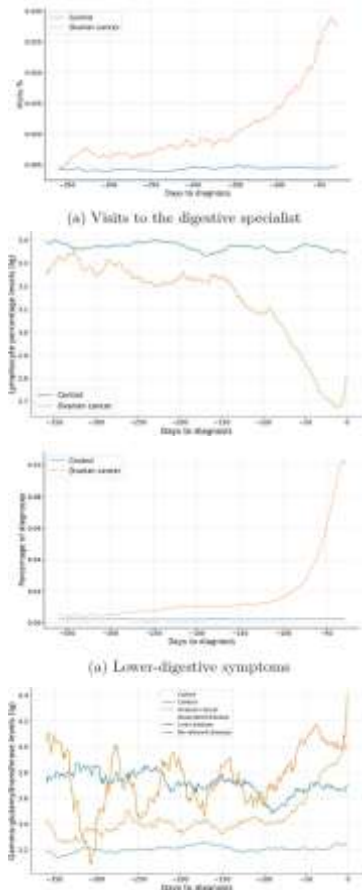
	Follow-up controls	Test controls	Test cases
N total per subclass	132468	7810	747
N positive predictions per subclass	8498	1842	418
Proportion of positive predictions per subclass	0.064	0.236	0.56
Proportion of positive predictions per class	0.061	0.011	0.56

Entrenamiento

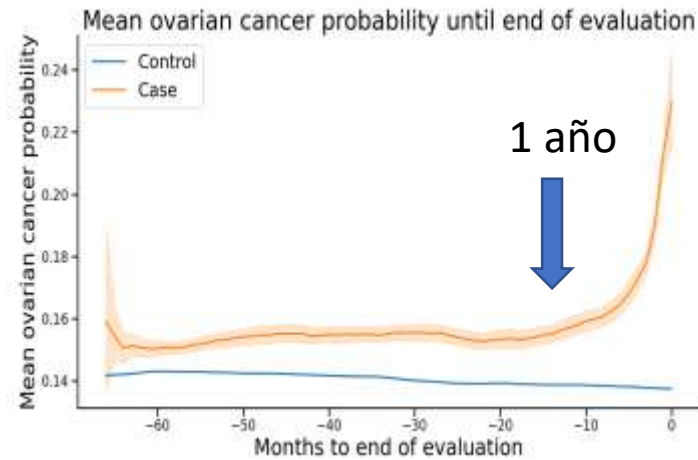
Los datos de la historia clínica permiten enriquecer la población de riesgo antes del diagnóstico

IA Explicable:

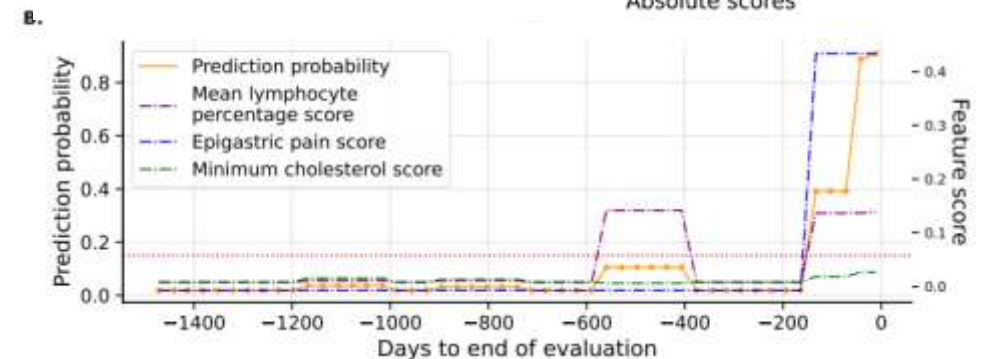
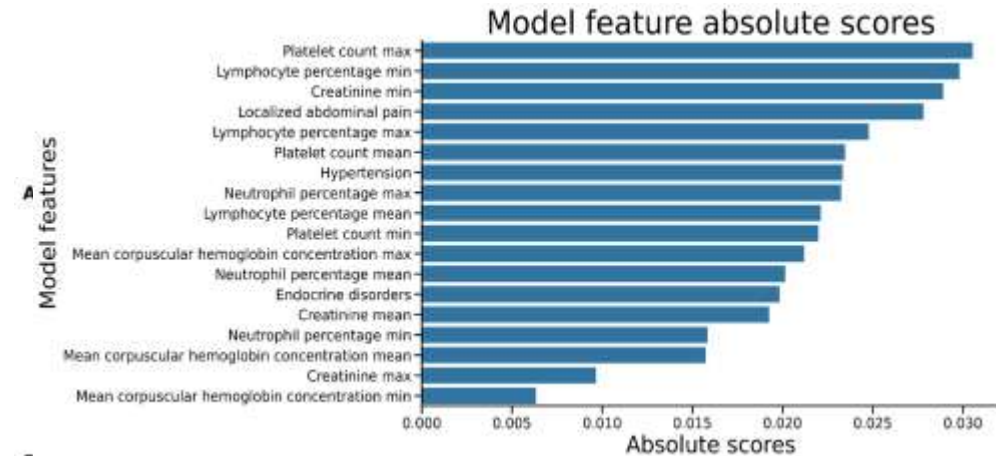
Las variables seleccionadas y usadas por el modelo son conocidas y resultan coherentes con el cáncer de ovario.



A partir de analíticas, diagnósticos, visitas al sistema de salud, etc., que por sí solas son poco específicas...



... se consigue una buena predicción a un año usando *Explainable Boosting Machine* (EBM), ajustando las ventanas de observación y los parámetros



Explicabilidad (XAI)

Necesidad de transparencia para:

- Entender qué variables contribuyen a una predicción.
- Trazabilidad de las decisiones
- Detectar "atajos" espurios (ej. usar un artefacto de imagen) o sesgos.
- Facilitar la confianza y aceptación por parte de clínicos y pacientes



La explicabilidad permite identificar sesgos y favorece la equidad

Si los datos reflejan desigualdades estructurales (ej. menor acceso a pruebas), los modelos pueden **aprender y amplificar** estos sesgos.



Riesgos:

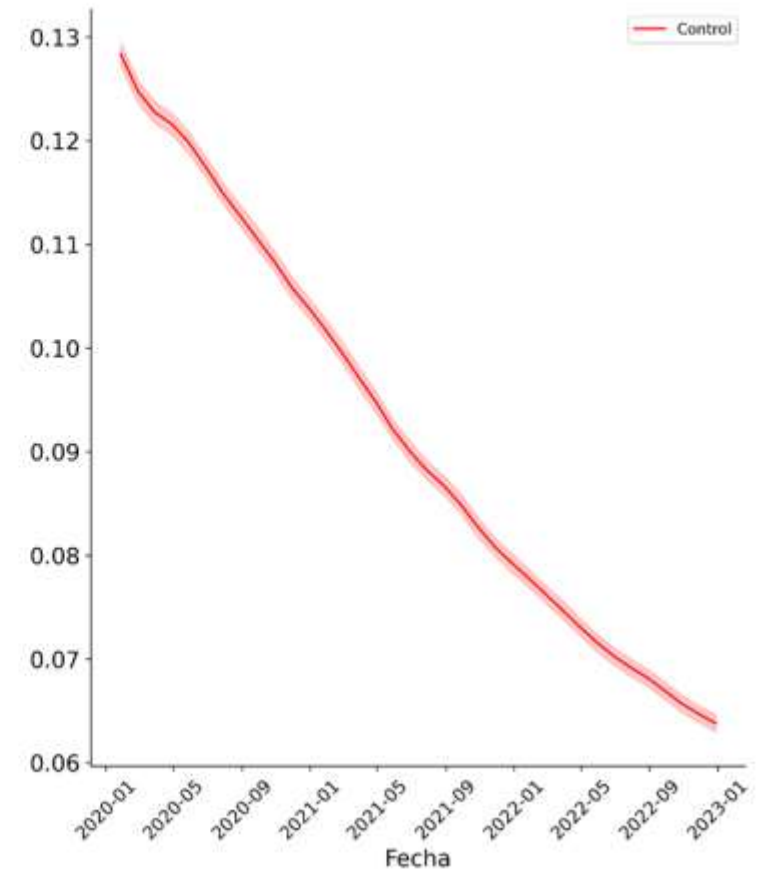
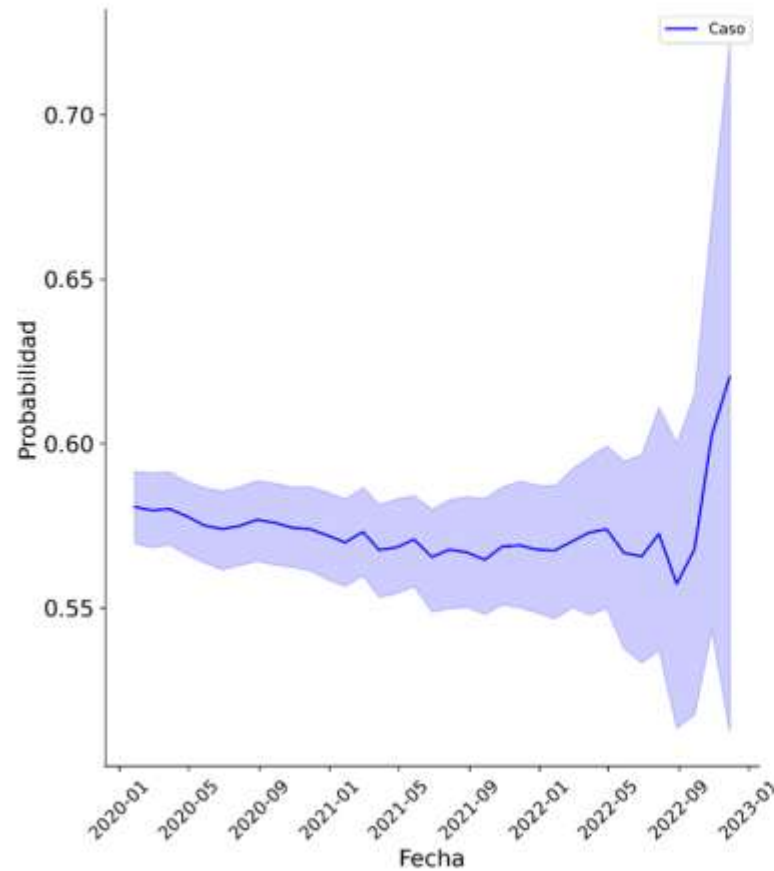
Modelos que perpetúan discriminación y funcionan peor en

- Minorías
- Mujeres
- Desfavorecidos
- Geográficamente deslocalizados

Asignación de probabilidades a las muestras

El modelo usa un umbral de probabilidad de 0.5 para decidir si una muestra es un caso (>0.5) o un control (<0.5).

Según mejora la información mejoran las predicciones



Resultados del predictor temprano de cáncer de ovario



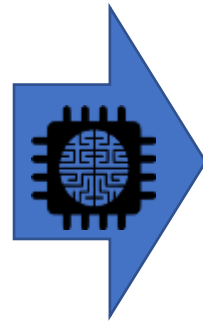
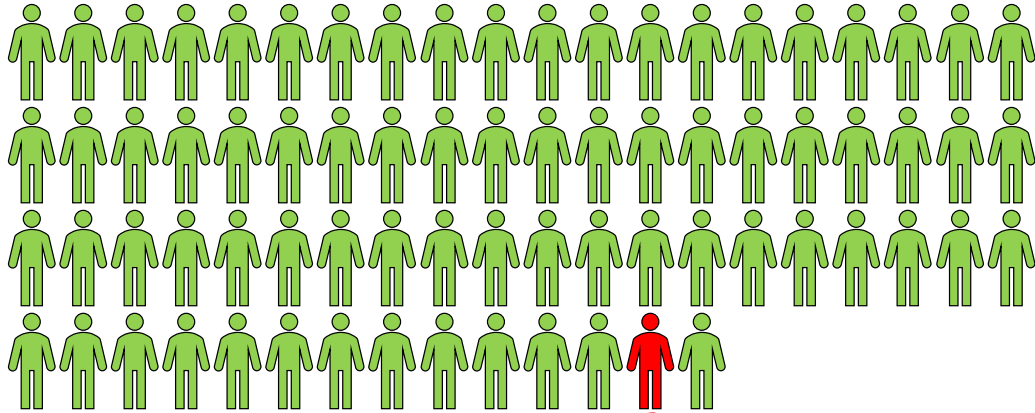
En segunda ronda de
revisión en *BMC Medical
Informatics and Decision
Making*



This study presents the development of an early prediction model for high-grade serous ovarian cancer (HGSOC) using real-world data from the Andalusian Health Population Database (BPS), containing electronic health records (EHR) of over 15 million patients. Leveraging the extensive data availability, the model aims to identify individuals at high risk of HGSOC without the need for specific tumor markers or prior stratification into risk groups. Utilizing an Explainable Boosting Machine (EBM) algorithm, the model incorporates diverse clinical variables including demographics, chronic diseases, symptoms, blood test results, and healthcare utilization patterns. **The model was trained and validated using a total of 3,088 HGSOC patients diagnosed between 2018 and 2022 along with 114,942 controls** of similar characteristics, to emulate the prevalence of the disease, achieving a **sensitivity of 0.65** and a **specificity of 0.85**. This study underscores the importance of using patient data from the general population, demonstrating that effective early detection models can be developed from routinely collected healthcare data. The approach addresses limitations of traditional screening methods by providing a **cost-effective and broadly applicable tool for early cancer detection**, potentially improving patient outcomes through timely interventions. The interpretability of the early prediction model also offers insights into the most significant predictors of cancer risk, further enhancing its utility in clinical settings.

Entendiendo los resultados del predictor

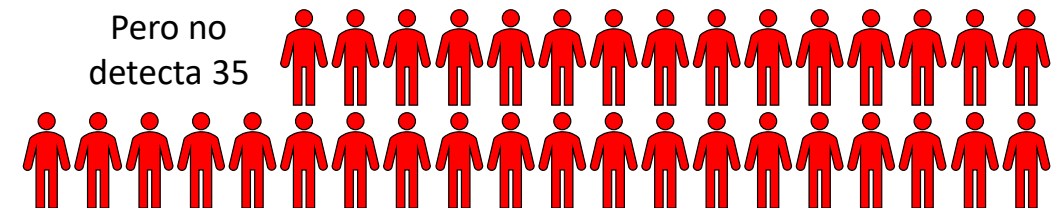
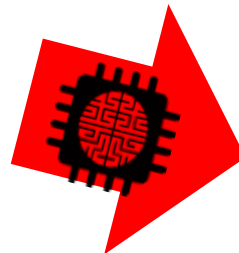
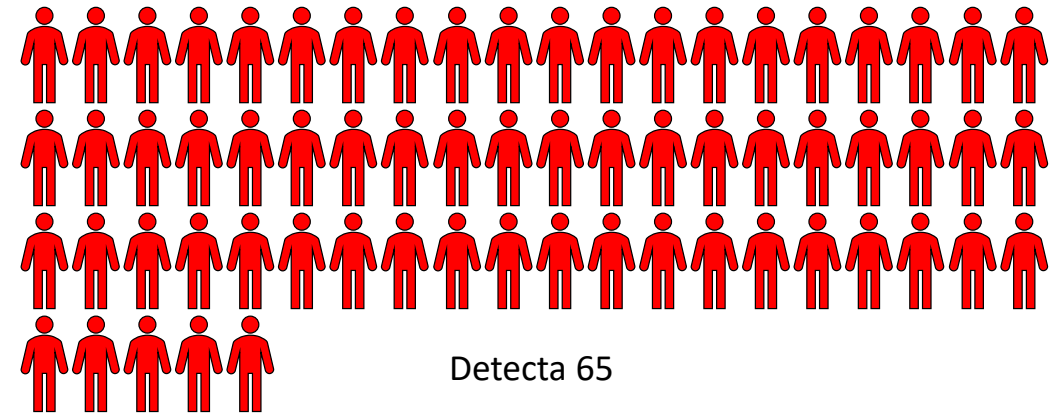
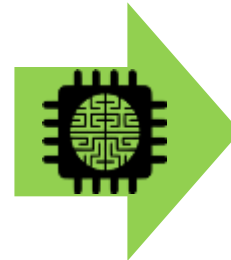
Prevalencia de cáncer de ovario 1 en 78. En un cribado etario convencional habría que cribar 78 mujeres para encontrar una positiva



Ningún cribado o predictor detecta todos los casos, la utilidad reside en aumentar la probabilidad pretest y hacer viable una intervención dirigida.



Después de aplicar el predictor, el 8.4% son verdaderos positivos. Habría que cribar 12 mujeres para encontrar una positiva



De cada 100...

Comparación con otros estudios

Model	Study size	Total OC	Sensitivity	Specificity	OC markers	Prediction	Population
Our model	118,030	3,088	0.65	0.85	No	OC 1 year	general
MCDM [1]	10,992	860	0.65	0.86	Yes	OC/non OC	OC risk
RNN [2]	37,293	88	0.47	0.90	Yes	OC 1 year	OC risk
SVN, KNN [3]	349	349	0.97	0.97	Yes	OC/benign OC	OC patients
RF, GBM, LGBM [4]	349	349	0.88	0.97	Yes	OC/benign OC	OC patients

1. Cai G, Huang F, Gao Y, Li X, Chi J, Xie J, Zhou L, Feng Y, Huang H, Deng T: **Artificial intelligence-based models enabling accurate diagnosis of ovarian cancer using laboratory tests in China: a multicentre, retrospective cohort study.** *The Lancet Digital health* 2024, **6**(3):e176-e186.
2. Abrego L, Zaikin A, Marino IP, Krivonosov MI, Jacobs I, Menon U, Gentry-Maharaj A, Blyuss O: **Bayesian and deep-learning models applied to the early detection of ovarian cancer using multiple longitudinal biomarkers.** *Cancer Medicine* 2024, **13**(7):e7163.
3. Juwono FH, Wong WK, Pek HT, Sivakumar S, Acula DD: **Ovarian cancer detection using optimized machine learning models with adaptive differential evolution.** *Biomedical Signal Processing and Control* 2022, **77**:103785.
4. Ahamad MM, Aktar S, Uddin MJ, Rahman T, Alyami SA, Al-Ashhab S, Akhdar HF, Azad A, Moni MA: **Early-stage detection of ovarian cancer based on clinical data using machine learning approaches.** *Journal of personalized medicine* 2022, **12**(8):1211.

Orden de magnitud comparativo: qué rendimiento necesitaría tener un predictor para ser útil en prevención

Cribado de cáncer colorrectal por detección de sangre oculta en heces:
Precisión (sensibilidad) aproximada 80%¹ con una participación del 42%²
Participación de Navarra, País Vasco 75% (detección 60%)

Detección real = 34% de la población objetivo

Cribado de cáncer de mama (con cita):
Precisión aproximada 80% [76-98%]³ con una participación del 74%⁴

Detección real = 59% de la población objetivo

La cita incrementa la participación probablemente por mayor conciencia del riesgo.

Cribado teórico de cáncer de ovario:
Precisión 65%, participación teórica 74%

Detección real = 48% de la población objetivo

1 <https://www.elsevier.es/es-revista-gastroenterologia-hepatologia-14-articulo-test-quimico-o-test-inmunologico-S0210570509003987>

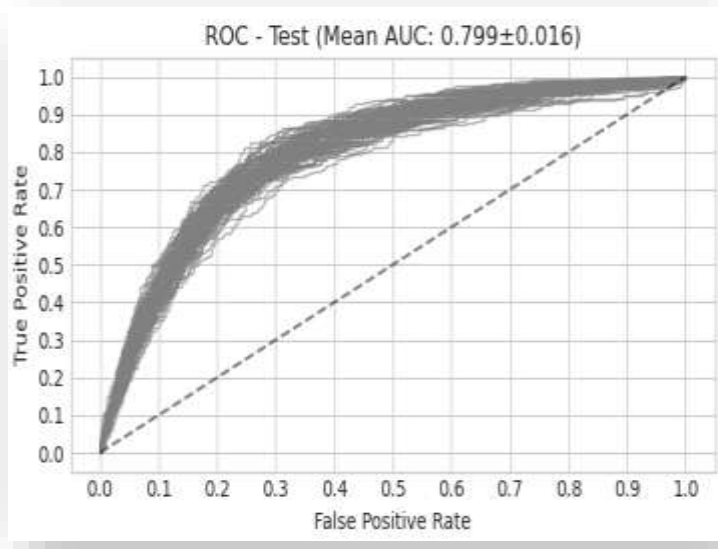
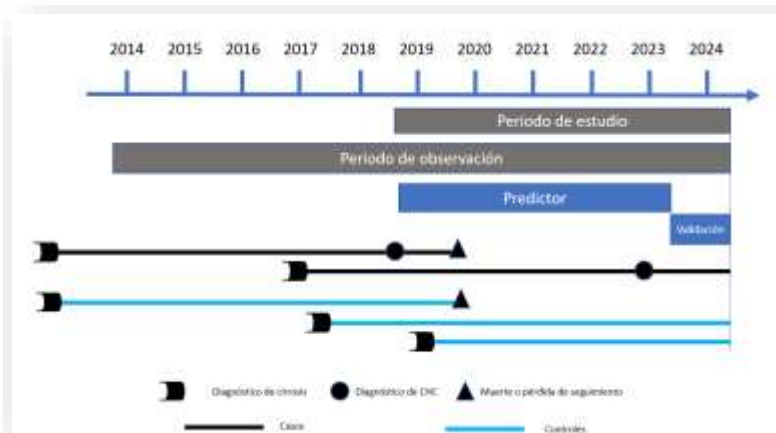
2 <https://isanidad.com/297490/la-participacion-media-en-los-programas-de-cribado-en-cancer-colorrectal- apenas-alcanza-el-42-con-importantes-desigualdades-entre-comunidades/>

3 https://saludextremadura.ses.es/areasaludcoria/wp-content/uploads/2024/07/DSAL_Programa-deteccion-cancer-de-mama-SES.pdf

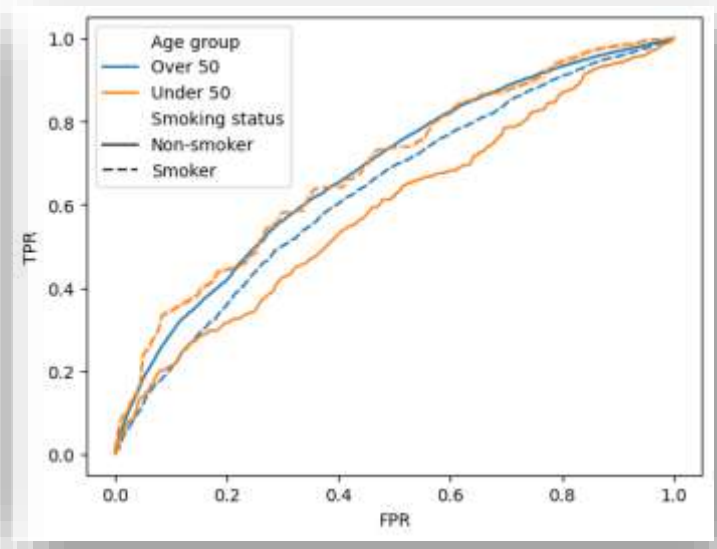
4 https://www.sanidad.gob.es/areas/calidadAsistencial/estrategias/cancer/docs/Espana_perfiles_nacionales_cancer_2023_ES.pdf



Cohorte retrospectiva de pacientes con cáncer de hígado junto con sus controles correspondientes, pacientes con enfermedad hepática crónica (cirrosis). Un total de **27.729** pacientes con cirrosis, con **1.082** casos de hepatocarcinoma (prevalencia del 3,9 %).



Hepatocarcinoma en pacientes cirróticos



Riesgo de cáncer de pulmón en la población general

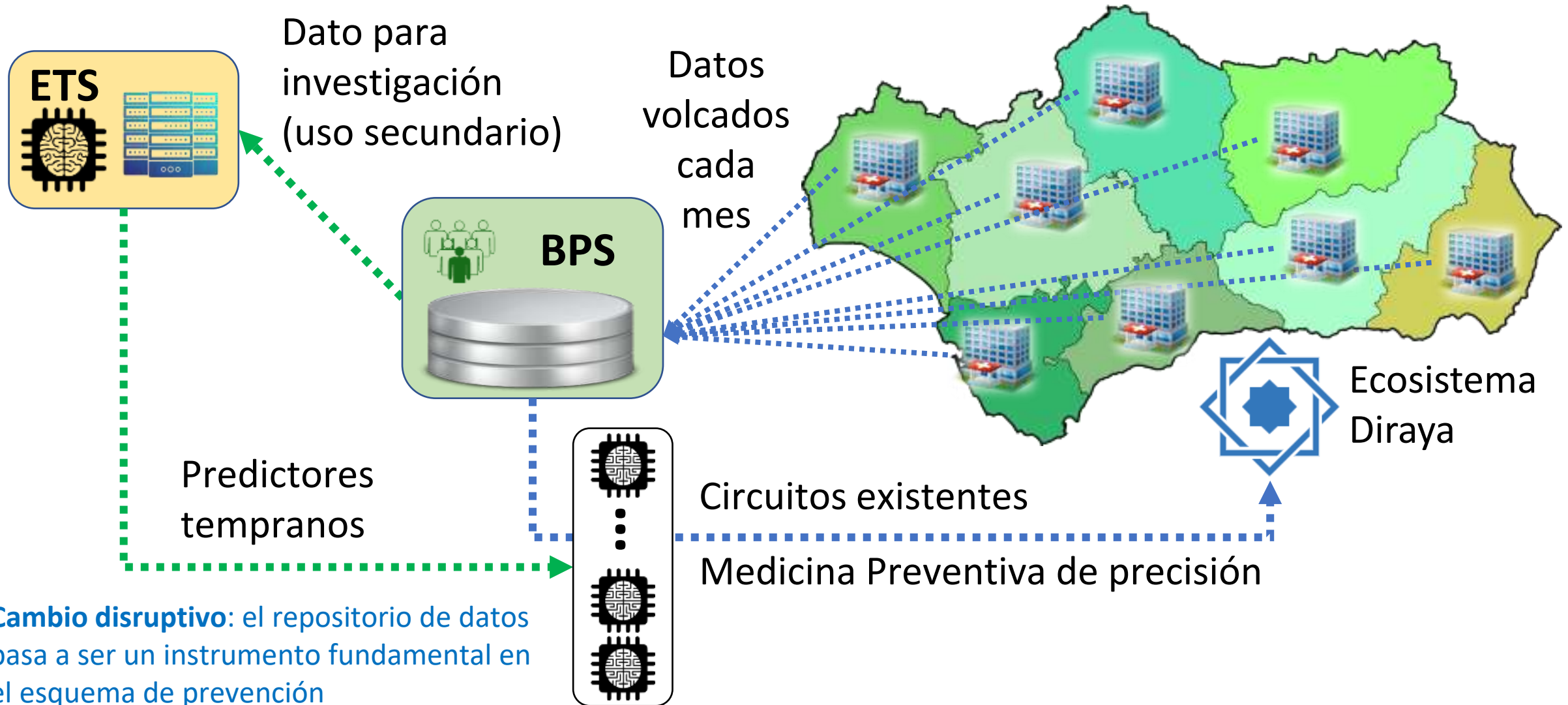
Predicción secundaria en población de riesgo:

- Altamente controlada
- Datos detallados
- Datos obtenidos regularmente
- Datos relevantes para la enfermedad

Prevención primaria en la población general:

- No se monitoriza
- Hay pocos datos disponibles (o ninguno)
- Hay menos datos disponibles sobre jóvenes
- Datos registrados de forma irregular
- No necesariamente relevantes para la enfermedad

Propuesta de implementación de predictores tempranos (Medicina Preventiva de Precisión)



Cambio disruptivo: el repositorio de datos pasa a ser un instrumento fundamental en el esquema de prevención

Condiciones para llevarlo a la práctica



1. Validación externa

Comprobar el rendimiento en otros sistemas sanitarios y poblaciones



2. Evaluación prospectiva

Medir su utilidad e impacto clínico en condiciones reales



3. Circuito asistencial claro

Definir qué hacer con una paciente de alto riesgo



4. Evaluación de equidad

Evitar sesgos por edad, territorio, nivel socioeconómico o uso del sistema



5. Integración segura

Desplegarlo en entornos seguros (EEDS)

Además de un buen modelo, para generar un impacto real hacen falta validación, circuito clínico, equidad y gobernanza

Conclusión: los datos asistenciales (RWD) son claves para la prevención de precisión

1 Datos clínicos rutinarios

El sistema sanitario ya registra señales longitudinales antes del diagnóstico: síntomas, visitas, analíticas, tratamientos y trayectorias.

2 Evidencia del mundo real

El uso secundario, con base ética, siguiendo el EEDS y con trazabilidad, permite transformar RWD en RWE robusta y reproducible.

3 Prevención anticipatoria

Los modelos explicables pueden seleccionar mejor a quién, cuándo y cómo intervenir antes de que la enfermedad sea visible.

La prevención de precisión tiene una capacidad transformadora del sistema de salud mucho mayor que cualquier otro aspecto de la medicina de precisión, y será viable cuando los datos que ya genera la asistencia se puedan (y quieran) reutilizar de forma segura para identificar riesgo, priorizar intervenciones y reducir enfermedad antes del diagnóstico.

Conclusión (2): los datos como infraestructura estratégica



El algoritmo debe ir a los datos y no los datos al algoritmo. Los datos no tienen por qué salir del sistema.

Mi agradecimiento a colaboradores, patrocinadores y agencias financiadoras



IA + datos del mundo real para reposicionar fármacos en COVID-19

De la predicción mecanística a la validación clínica

- LA IDEA**
Objetivo: identificar medicamentos ya aprobados que podrían tener un segundo uso terapéutico frente a la COVID-19.
Estrategia: combinar predicción computacional e investigación con datos clínicos reales.
- ESTUDIO 1: MODELO MATEMÁTICO + IA**
Mapa de la enfermedad: Se modela el mecanismo biológico de la COVID-19 desde el virus hasta el sistema de defensa.
Modelo mecanístico: A partir de datos de expresión génica se extrae el circuito de regulación por fármacos.
Aprendizaje automático: Se construye una red neuronal que relaciona fármacos terapéuticos con los circuitos de regulación.
Interpretabilidad: SHAP ayuda a identificar qué circuitos y fármacos influyen más en la predicción de la enfermedad.
Resultado: se priorizan cientos de candidatos a reposicionando 870 fármacos, 286 fármacos.
- ¿QUÉ PREDICE EL MODELO?**
No basta con fármacos autorizados, también se pueden usar fármacos que costaban más.
No basta con fármacos autorizados, también se pueden usar fármacos que costaban más.
La IA genera hipótesis biológicamente razonables, no detectar correlaciones.
Ejemplos: Aspirina, Paracetamol, Vitamina D, Zinc, Vitamina C, Vitamina E, Vitamina K, Vitamina B12, Vitamina B6, Vitamina B1, Vitamina B2, Vitamina B3, Vitamina B5, Vitamina B7, Vitamina B9, Vitamina B10, Vitamina B11, Vitamina B12, Vitamina B13, Vitamina B14, Vitamina B15, Vitamina B16, Vitamina B17, Vitamina B18, Vitamina B19, Vitamina B20, Vitamina B21, Vitamina B22, Vitamina B23, Vitamina B24, Vitamina B25, Vitamina B26, Vitamina B27, Vitamina B28, Vitamina B29, Vitamina B30, Vitamina B31, Vitamina B32, Vitamina B33, Vitamina B34, Vitamina B35, Vitamina B36, Vitamina B37, Vitamina B38, Vitamina B39, Vitamina B40, Vitamina B41, Vitamina B42, Vitamina B43, Vitamina B44, Vitamina B45, Vitamina B46, Vitamina B47, Vitamina B48, Vitamina B49, Vitamina B50, Vitamina B51, Vitamina B52, Vitamina B53, Vitamina B54, Vitamina B55, Vitamina B56, Vitamina B57, Vitamina B58, Vitamina B59, Vitamina B60, Vitamina B61, Vitamina B62, Vitamina B63, Vitamina B64, Vitamina B65, Vitamina B66, Vitamina B67, Vitamina B68, Vitamina B69, Vitamina B70, Vitamina B71, Vitamina B72, Vitamina B73, Vitamina B74, Vitamina B75, Vitamina B76, Vitamina B77, Vitamina B78, Vitamina B79, Vitamina B80, Vitamina B81, Vitamina B82, Vitamina B83, Vitamina B84, Vitamina B85, Vitamina B86, Vitamina B87, Vitamina B88, Vitamina B89, Vitamina B90, Vitamina B91, Vitamina B92, Vitamina B93, Vitamina B94, Vitamina B95, Vitamina B96, Vitamina B97, Vitamina B98, Vitamina B99, Vitamina B100.
- ESTUDIO 2: VALIDACIÓN CON DATOS DEL MUNDO REAL**
Se analiza la efectividad de los fármacos en pacientes con COVID-19.
Se valida el modelo matemático con datos del mundo real.
Resultado: 31 fármacos muestran efectos antiinflamatorios significativos.
- EJEMPLOS DE FÁRMACOS PROTECTORES**
Ejemplos: Aspirina, Paracetamol, Vitamina D, Zinc, Vitamina C, Vitamina E, Vitamina K, Vitamina B12, Vitamina B6, Vitamina B1, Vitamina B2, Vitamina B3, Vitamina B5, Vitamina B7, Vitamina B9, Vitamina B10, Vitamina B11, Vitamina B12, Vitamina B13, Vitamina B14, Vitamina B15, Vitamina B16, Vitamina B17, Vitamina B18, Vitamina B19, Vitamina B20, Vitamina B21, Vitamina B22, Vitamina B23, Vitamina B24, Vitamina B25, Vitamina B26, Vitamina B27, Vitamina B28, Vitamina B29, Vitamina B30, Vitamina B31, Vitamina B32, Vitamina B33, Vitamina B34, Vitamina B35, Vitamina B36, Vitamina B37, Vitamina B38, Vitamina B39, Vitamina B40, Vitamina B41, Vitamina B42, Vitamina B43, Vitamina B44, Vitamina B45, Vitamina B46, Vitamina B47, Vitamina B48, Vitamina B49, Vitamina B50, Vitamina B51, Vitamina B52, Vitamina B53, Vitamina B54, Vitamina B55, Vitamina B56, Vitamina B57, Vitamina B58, Vitamina B59, Vitamina B60, Vitamina B61, Vitamina B62, Vitamina B63, Vitamina B64, Vitamina B65, Vitamina B66, Vitamina B67, Vitamina B68, Vitamina B69, Vitamina B70, Vitamina B71, Vitamina B72, Vitamina B73, Vitamina B74, Vitamina B75, Vitamina B76, Vitamina B77, Vitamina B78, Vitamina B79, Vitamina B80, Vitamina B81, Vitamina B82, Vitamina B83, Vitamina B84, Vitamina B85, Vitamina B86, Vitamina B87, Vitamina B88, Vitamina B89, Vitamina B90, Vitamina B91, Vitamina B92, Vitamina B93, Vitamina B94, Vitamina B95, Vitamina B96, Vitamina B97, Vitamina B98, Vitamina B99, Vitamina B100.
- ¿POR QUÉ ES INNOVADOR?**
La innovación no es solo predecir, sino cerrar el ciclo entre mecanismo, predicción y evidencia clínica real.



El Genoma Español: Claves para una Medicina de Precisión

EL DESAFÍO: UNA VARIABILIDAD GENÉTICA ÚNICA

Casi 1/3 de las variantes genéticas encontradas eran nuevas.

Un estudio de 287 familias españolas reveló miles de polimorfismos únicos de esta población.

Diferente perfil de riesgo para enfermedades raras. Las frecuencias de variantes de enfermedades mendelianas (ej. distrofia de retina) difieren notablemente de otras poblaciones.

Sin datos locales, una variante puede ser malinterpretada. El uso de un catálogo local reduce 3 veces más los falsos positivos en la búsqueda de genes causantes de enfermedades.

LA SOLUCIÓN: EL CATÁLOGO GENÉTICO CSVS

CSVS: Una base de datos con más de 2.000 genomas españoles. Creado mediante un esfuerzo colaborativo (Enfermedades) con apoyo de miles de proyectos genómicos.

Permite crear "poblaciones de control" a medida. Se pueden incluir subgrupos de enfermedades para reducir frecuencias específicas de una población particular.



El Atlas de Nuestro ADN

El Futuro de la Medicina Preventiva

Una historia sobre el Riesgo Poligénico Español



Joaquin Dopazo Director of the BioAI Data Platform, BBRC / CRG / HMRB / ... Barcelona, Catalonia Real Academia Nacional de Medicina

Promote this post to reach people who matter to you. Boost

I'm pleased to share our recent article: "Ethical and secure evidence generation from regionwide clinical data through a collaborative environment for advancing predictive care." Read it here: <https://link.in/dp1AXXQx>

Ensuring data protection is one of the main challenges in clinical research using sensitive patient information. In this work, we present how secure processing environments (SPEs) can enable the ethical and compliant secondary use of real-world data (RWD) for large-scale, impactful evidence generation. Our study describes the Andalusian Platform for Medical Evidence Generation (PAGEM) — a secure environment embedded within the Andalusian Health System — integrating data from over 15 million patients in the Andalusian Health Population Database (BPS). This infrastructure allows robust, GDPR-aligned research on treatment efficacy, survival, and predictive models, while anticipating the requirements of the European Health Data Space (EHDS).

By combining clinical RWD, ethic governance, and a comprehensive legal framework, we offer a scalable model for regionwide healthcare innovation. This approach not only enables predictive, cost-effective care — particularly for aging populations — but also reinforces the dual role of healthcare systems as both providers of care and generators of knowledge and economic growth. This work represents a step forward in making health data securely actionable for scientific and societal benefit.

#RWD #RWE #HealthcareInnovation #DataGovernance #GDPR #EHDS #PredictiveMedicine #PublicHealth #Andalusia #DigitalHealth #Healthcare

Profile viewers: 551
Post impressions: 1,682

Ethical and secure evidence generation from regionwide clinical data through a collaborative environment for advancing predictive care




Gracias por su atención

 <https://www.linkedin.com/in/joaquindopa>

 <https://joaquindopazo.substack.com/>

 <https://www.instagram.com/datos.que.curan/>

 @xdopazo.bsky.social

 @xdopazo

