

Alumna: María Uriburu Gray  
Mentor: Joan Gasull Jolis

# Más que raras

## Exploración de las Enfermedades Raras en Europa y su impacto regional en España



# Agenda



¿Qué son las enfermedades raras?  
Definiciones necesarias



Objetivos del estudio



Colección de datos



Tratamiento de datos



Resultados



Conclusiones



# ¿Qué son las enfermedades raras?

## Enfermedad Rara

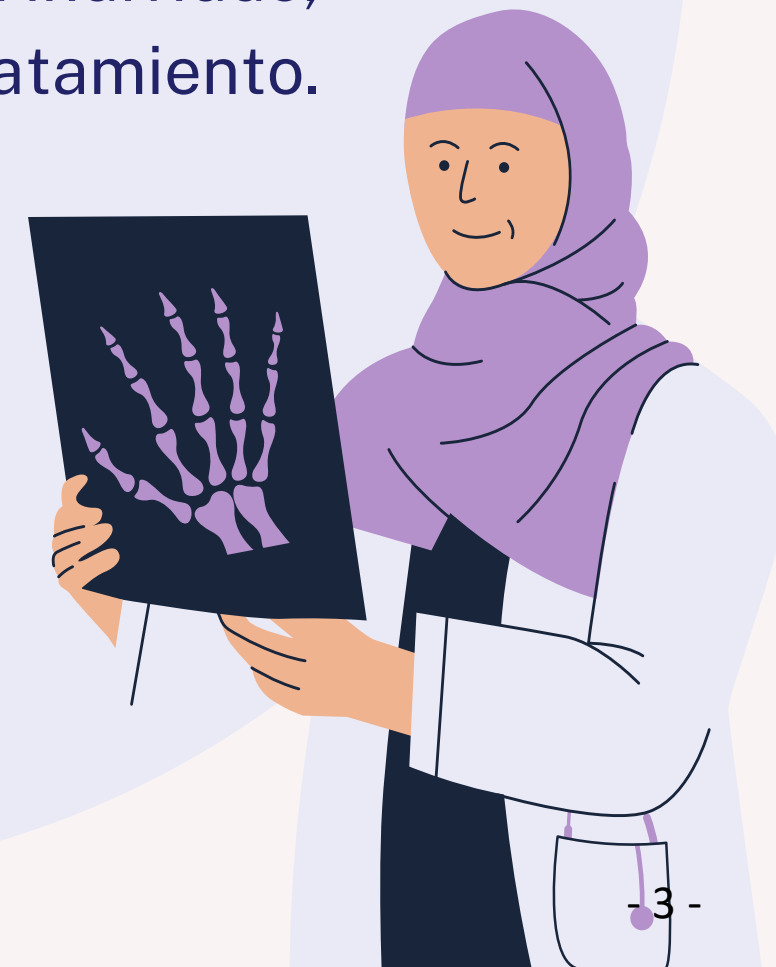
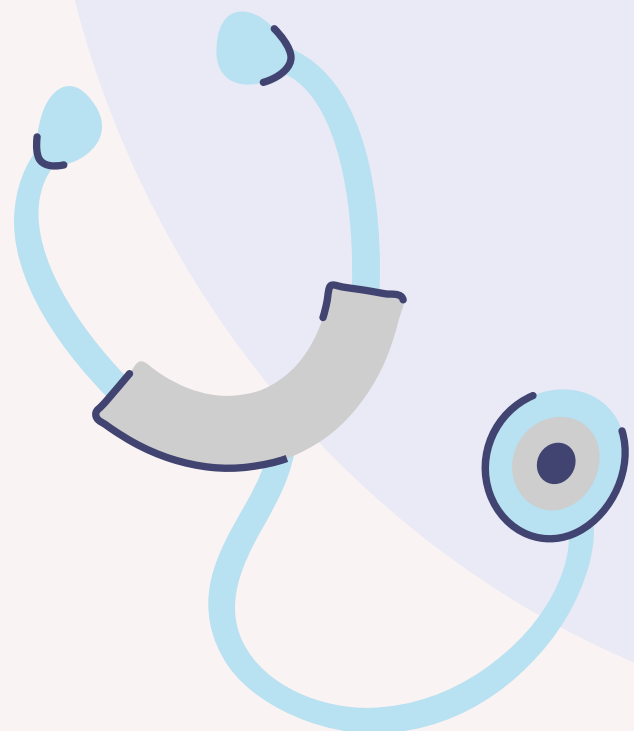
Aquellas que, con peligro de muerte o invalidez crónica, tiene una **prevalencia** menor de 5 casos por cada 10.000 habitantes.

## Fenotipos

Características observables de una enfermedad en un individuo, como signos, **síntomas** y respuestas al tratamiento.

## Incidencia vs Prevalencia

La **incidencia** mide el número de **nuevos casos** de una enfermedad en una población y tiempo determinado, mientras que la **prevalencia** mide el total de personas que **tienen** una enfermedad.



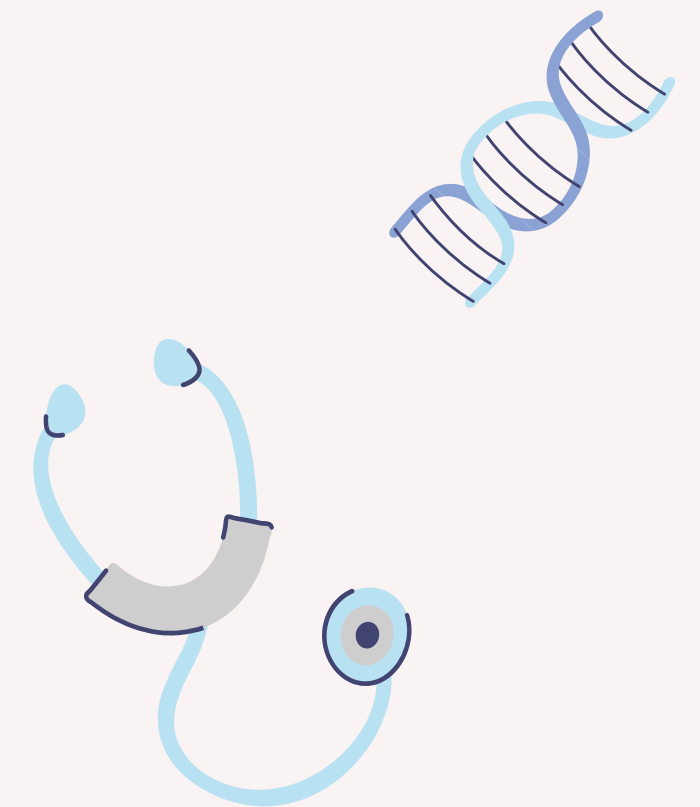
# Objetivos del estudio

**Identificar patrones que contribuyan a la investigación y gestión sanitaria de las Enfermedades Raras**

Estudio de la complejidad de las ER.  
¿Existe relación entre el número de genes implicados y el número de fenotipos que presentan?

Estudio regional de la carga de ER en España.  
¿Existe alguna relación entre la concentración de diagnósticos de ER y la cantidad de población?

Estudio de la afectación por género.  
¿Existe una mayor predisposición a padecer estas enfermedades según el género? En caso afirmativo, ¿cuál es más afectado: hombres o mujeres?



# Colección de Datos

## API PubMed

Acceso al registro de publicaciones científicas que contienen “rare disease” en el título y año de publicación.

## API Orphadata

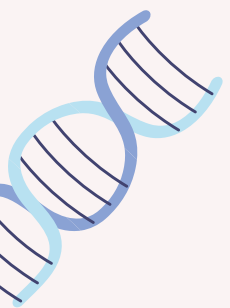
Acceso al registro completo de Enfermedades Raras. Contiene información de la epidemiología, historia clínica, prevalencia, genes y fenotipos.

## Informes ReeR

Informes del Registro Estatal de Enfermedades Raras con información de los casos notificados por cada comunidad autónoma.

## Datos INE

Población total por Comunidades Autónomas en 2018.

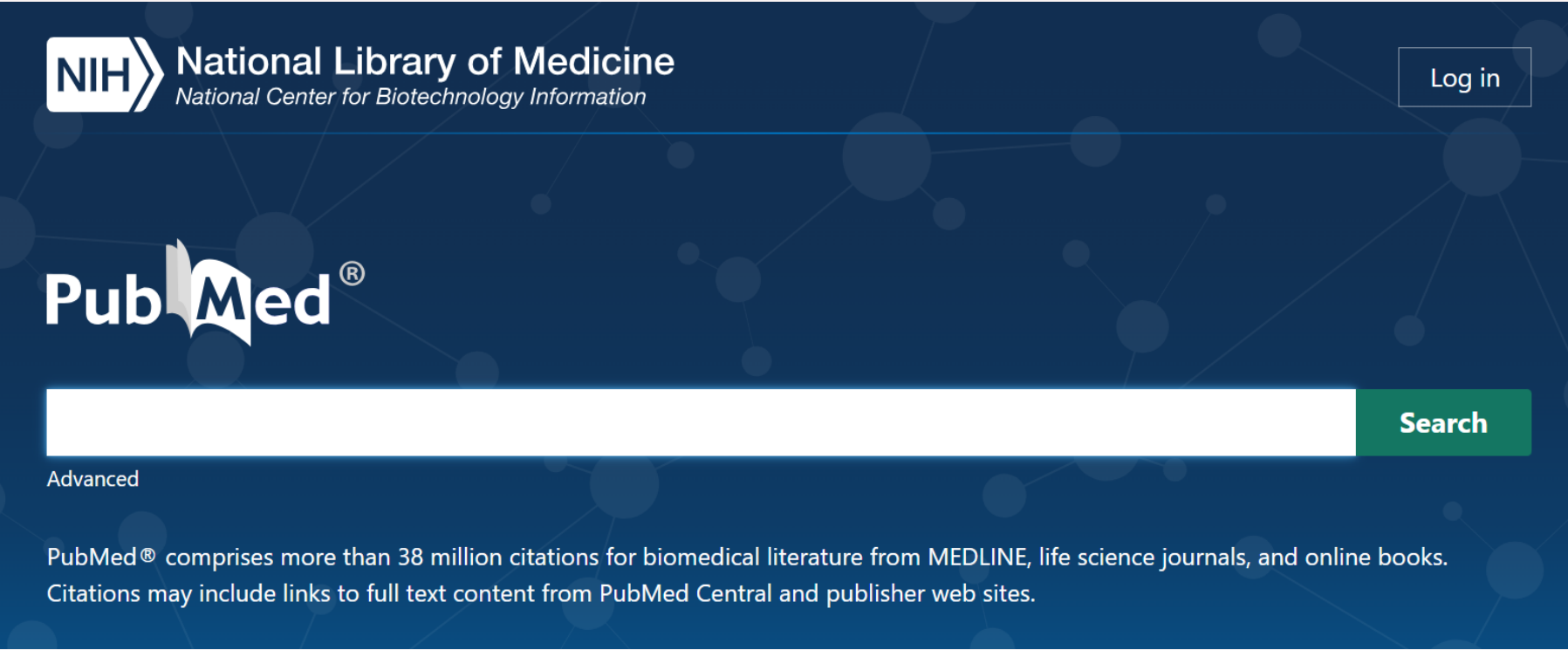




# Tratamiento de Datos

## API PubMed

Base de datos gratuita de la Biblioteca Nacional de Medicina de EE.UU.

- Base de datos muy limpia.
- 2617 filas y 5 columnas.
- Interés en el nº de publicaciones por año.



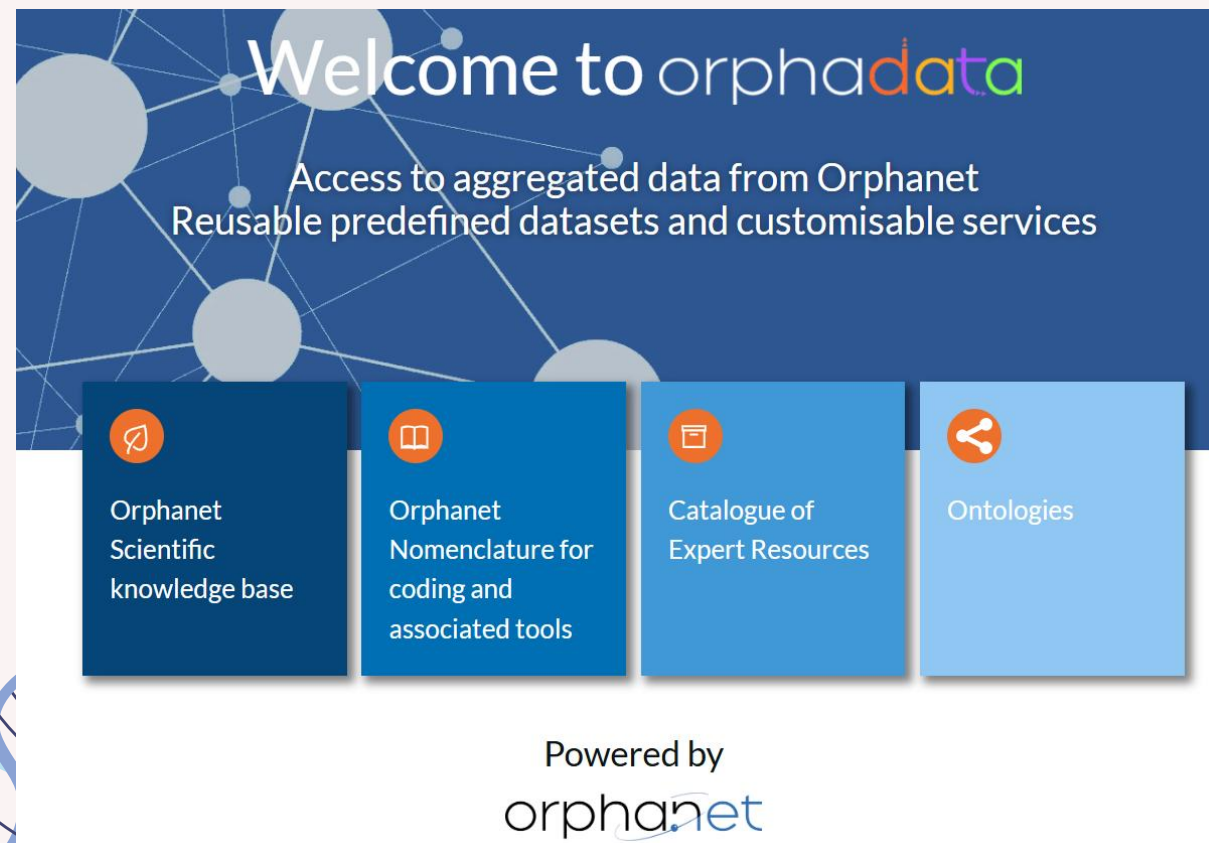
# pmid	# pub_year	# pubtype	# journal		
Missing: 0 (0%) Distinct: 2391 (91%)	Missing: 2 (<1%) Distinct: 2381 (91%)	Missing: 0 (0%) Distinct: 79 (3%)	Missing: 0 (0%) Distinct: 1184 (45%)		
 <div>2381 Distinct values</div>	 <div>1184 Distinct values</div>	<div>Journal Article 28% Case Reports; Journal Article 25% Journal Article; Research Su... 8% Other 39%</div>			
Min 44654 Max 40191800	Min 1909 Max 2025				
0	19973918	A Case of a Rare Disease of the Long	1909	Journal Article	Proceedings of the Royal Society of I
1	19979883	A Rare Disease in Two Brothers.	1917	Journal Article	Proceedings of the Royal Society of I
2	30438855	A Rare Disease of the Conjunctiva, w	1933	Journal Article	Glasgow medical journal
3	20282579	Sandifort's Observaciones, chapter I,	1946	Biography; Journal Article	Bulletin of the history of medicine
4	20985704	Chronic erosive, granulomatous, atro	1946	Journal Article	Gastroenterology
5	15410547	Is early osteochondritis of the patella	1950	Journal Article	Acta chirurgiae orthopaedicae et tra
6	14822046	[Rare disease of the cornea caused b	1951	Journal Article	Ceskoslovenska oftalmologie
7	13026929	[An unusual case of a rare disease, g	1952	Journal Article	La Presse medicale
8	13012938	[A rare disease caused by accident: M	1952	Journal Article	Minerva medica



# Tratamiento de Datos

## API Orphadata

Orphanet es una base de datos online de acceso libre dedicada a las enfermedades raras y los medicamentos huérfanos. Orphadata es una plataforma que proporciona datos a la comunidad científica.

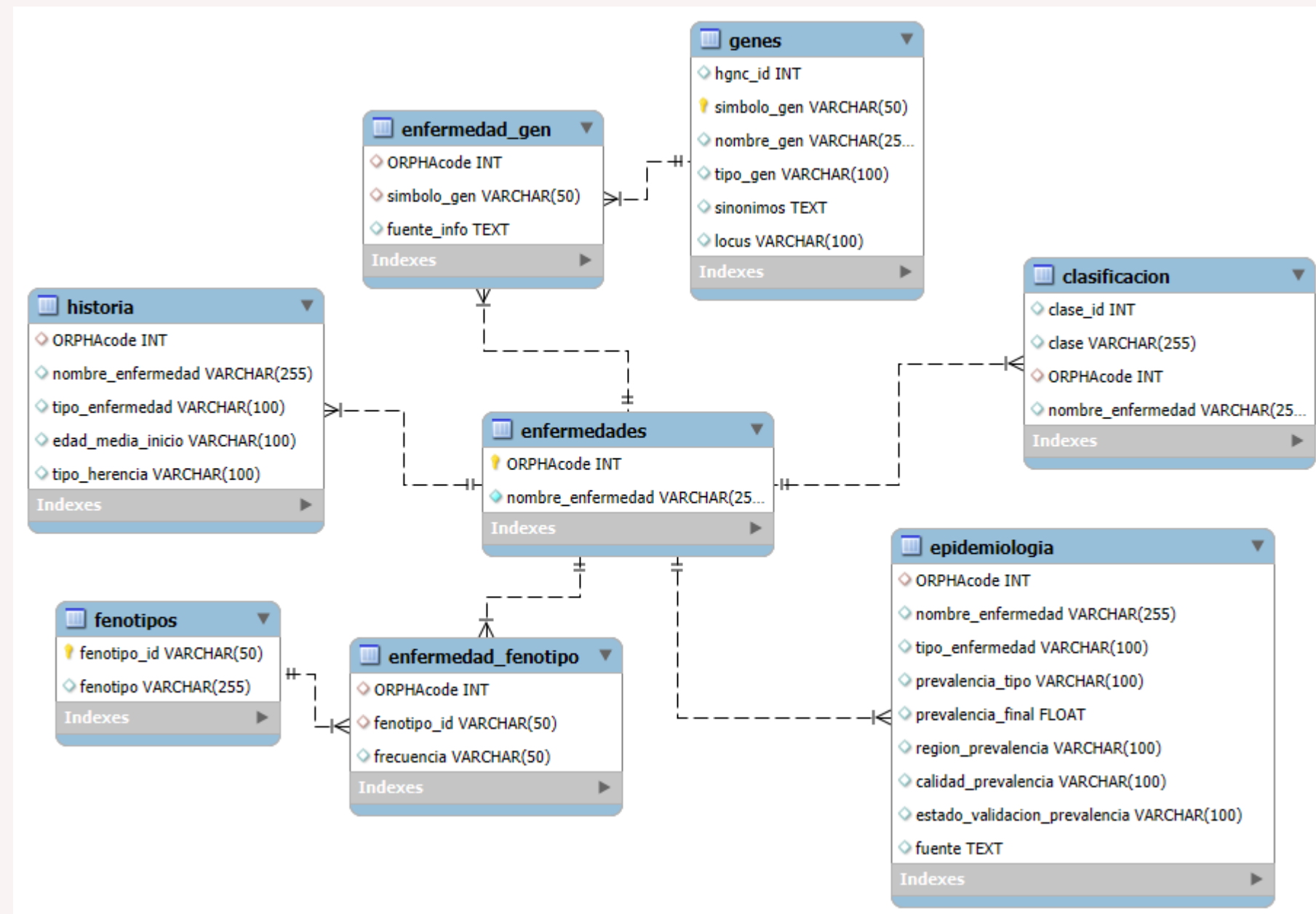


Welcome to orphadata

Access to aggregated data from Orphanet  
Reusable predefined datasets and customisable services

- Orphanet Scientific knowledge base
- Orphanet Nomenclature for coding and associated tools
- Catalogue of Expert Resources
- Ontologies

Powered by orphanet



- Reestructuración de tablas.
- Limpieza de datos.
- Tablas utilizadas: **ENFERMEDADES**, **GENES** y **FENOTIPOS**.

# Tratamiento de Datos

# ReeR

**El Registro Estatal de Enfermedades Raras** es un sistema de información del Ministerio de Sanidad, a través del Instituto de Salud Carlos III, **registra todos los casos de enfermedades raras en España.**

Red de registros autonómicos (RAER) transmiten los datos al ReeR central

**INE**

Instituto Nacional de Estadística, es el organismo encargado de producir las **estadísticas oficiales** de España.

- **PDFPlumber** es una librería de Python para extraer datos y tablas de archivos PDF.
- **Información limitada.** Faltan años, Comunidades Autónomas y faltan enfermedades.
- **Tablas utilizadas:**
  - **Casos notificados 2010 – 2018 por CC.AA.**
  - **Población total 2018 por CC.AA. (INE)**
  - **Casos según género 2020 – 2022.**
  - **Enfermedades según género 2022.**

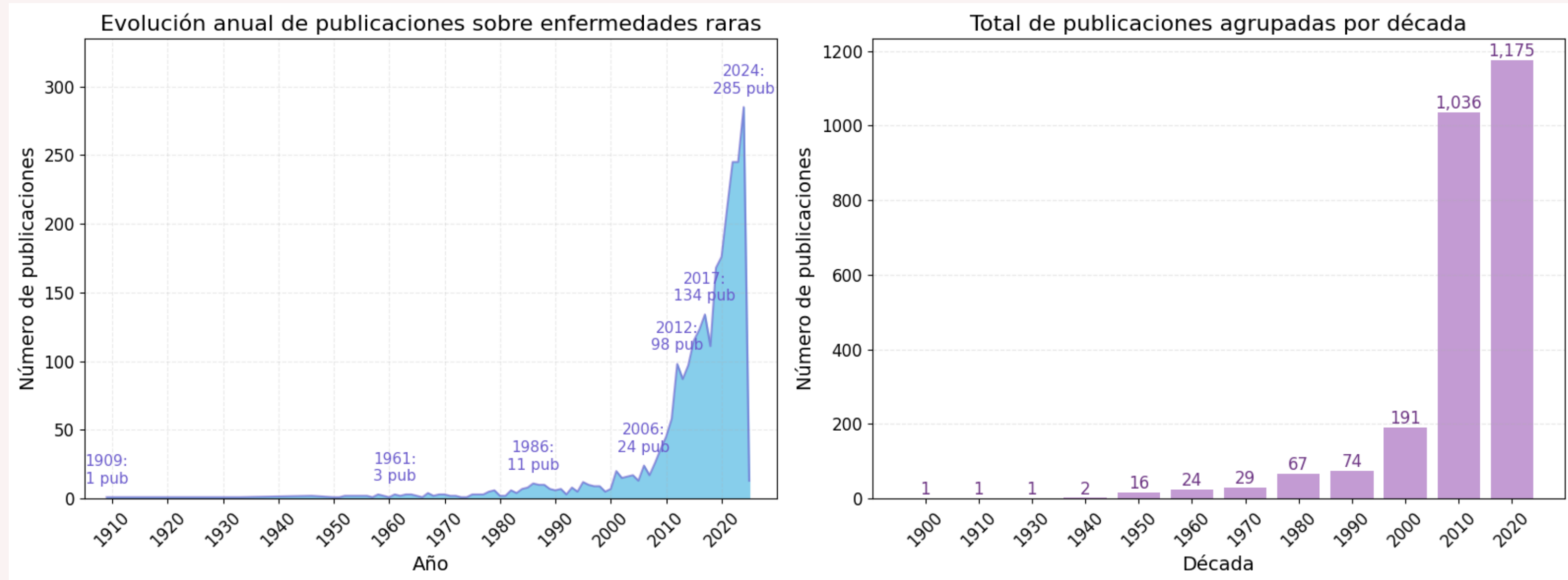




# Resultados

## – PubMed –

### Extracción del número de publicaciones científicas con “*rare disease*” en el título

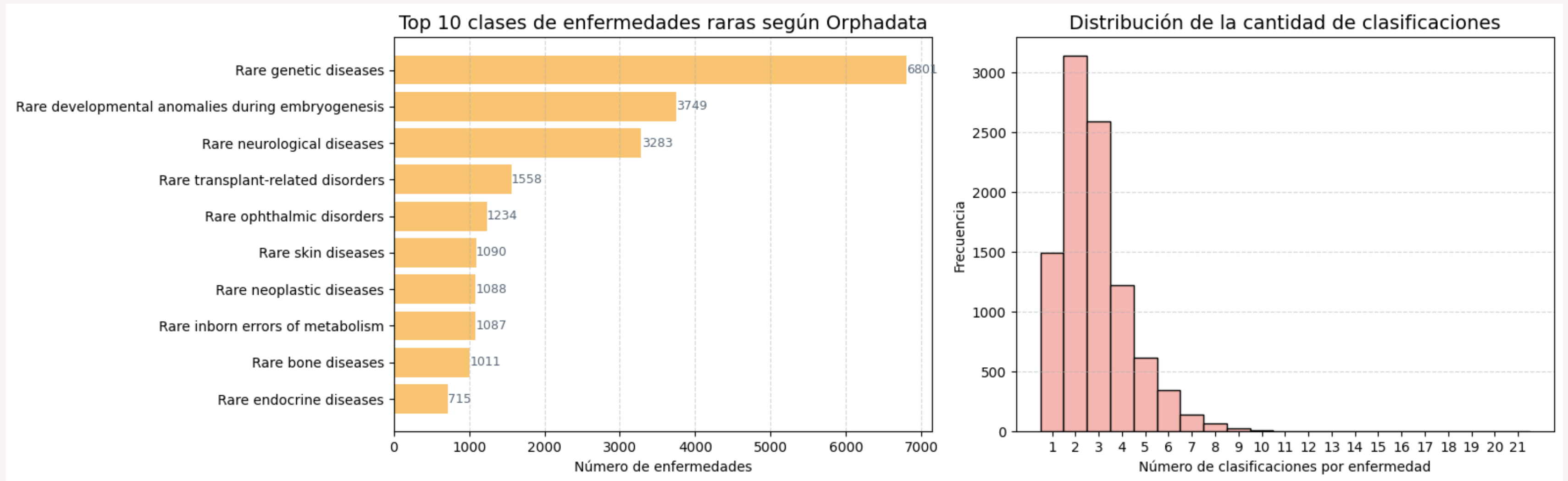


- ❑ La primera publicación de enfermedades raras fue en el año 1906.
- ❑ Aumento significativo del número de investigaciones publicadas en la década del 2010.

# Resultados

## - Orphadata -

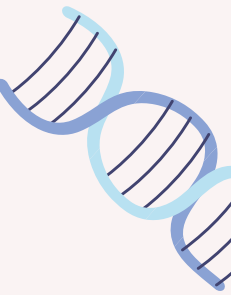
**9673 registros de Enfermedades Raras distribuidas en 35 clases según Orphadata**



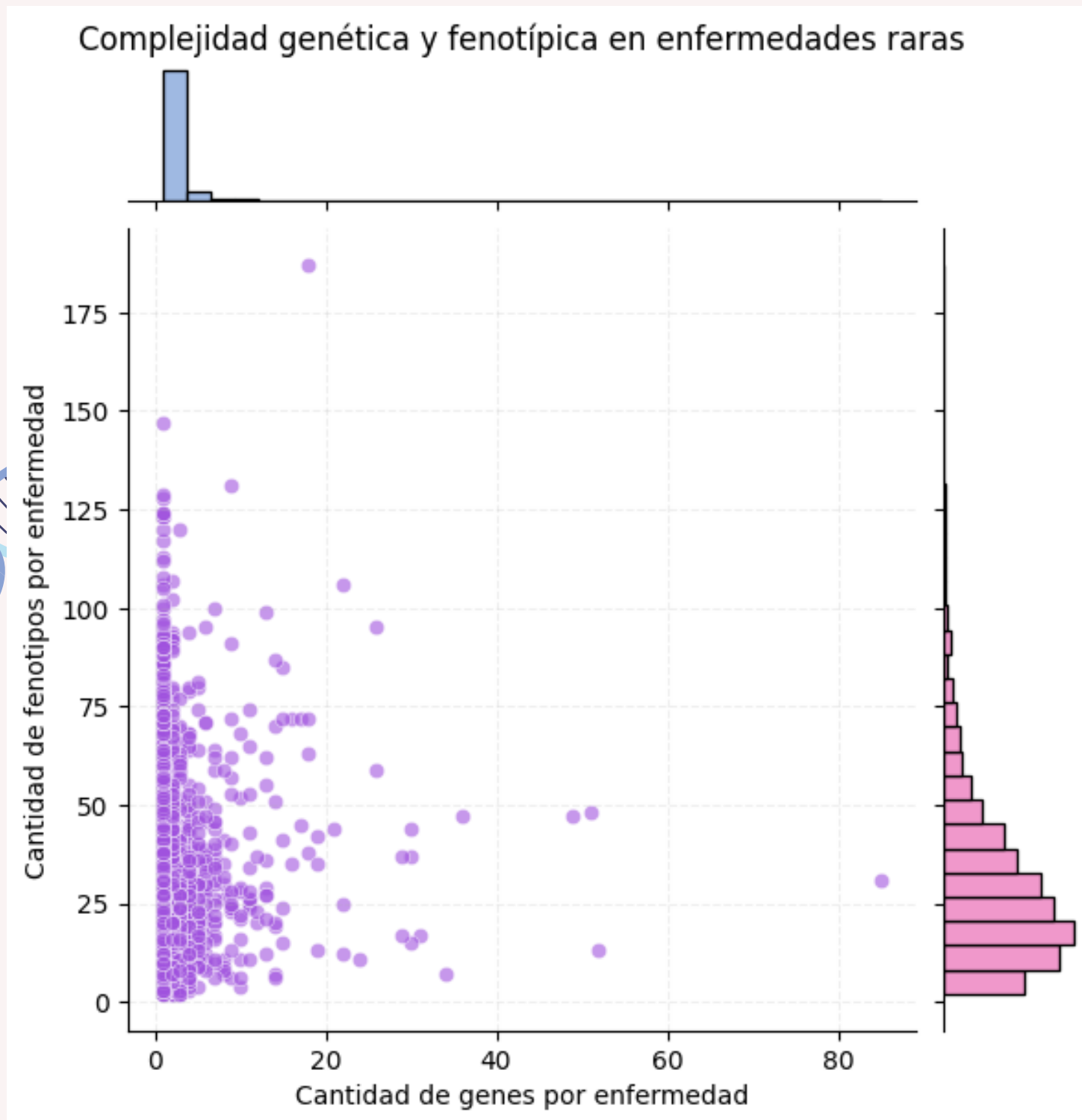
- ❑ La mayoría de las enfermedades tienen un origen genético.
- ❑ Las enfermedades pueden catalogarse en más de una clase, la mayoría pertenecen a 2 o 3 clases.

# Resultados

## - Orphadata -



### Análisis de la dependencia del número de genes y fenotipos asociados a una ER



❑ Correlación de Spearman = 0.0392 → No se observa.

❑ Genes por enfermedad:

Discapacidad intelectual no sindrómica (ORPHAcode: 528084) está asociada a **108 genes** distintos.

❑ Enfermedades por gen:

Genes **LMNA** y **TP53** implicados en **21 enfermedades** cada uno.

# Resultados

## - Ornhadata -

### ¿Sabías que...?

¡Es posible consultar información genética directamente desde Python!

Existe una librería llamada BioPython que permite acceder a bases de datos biomédicas como el NCBI (National Center for Biotechnology Information) y obtener información de genes, proteínas y secuencias genómicas directamente desde tu código.

```
1 from Bio import Entrez
2 Entrez.email = "A.N.Other@example.com"
```

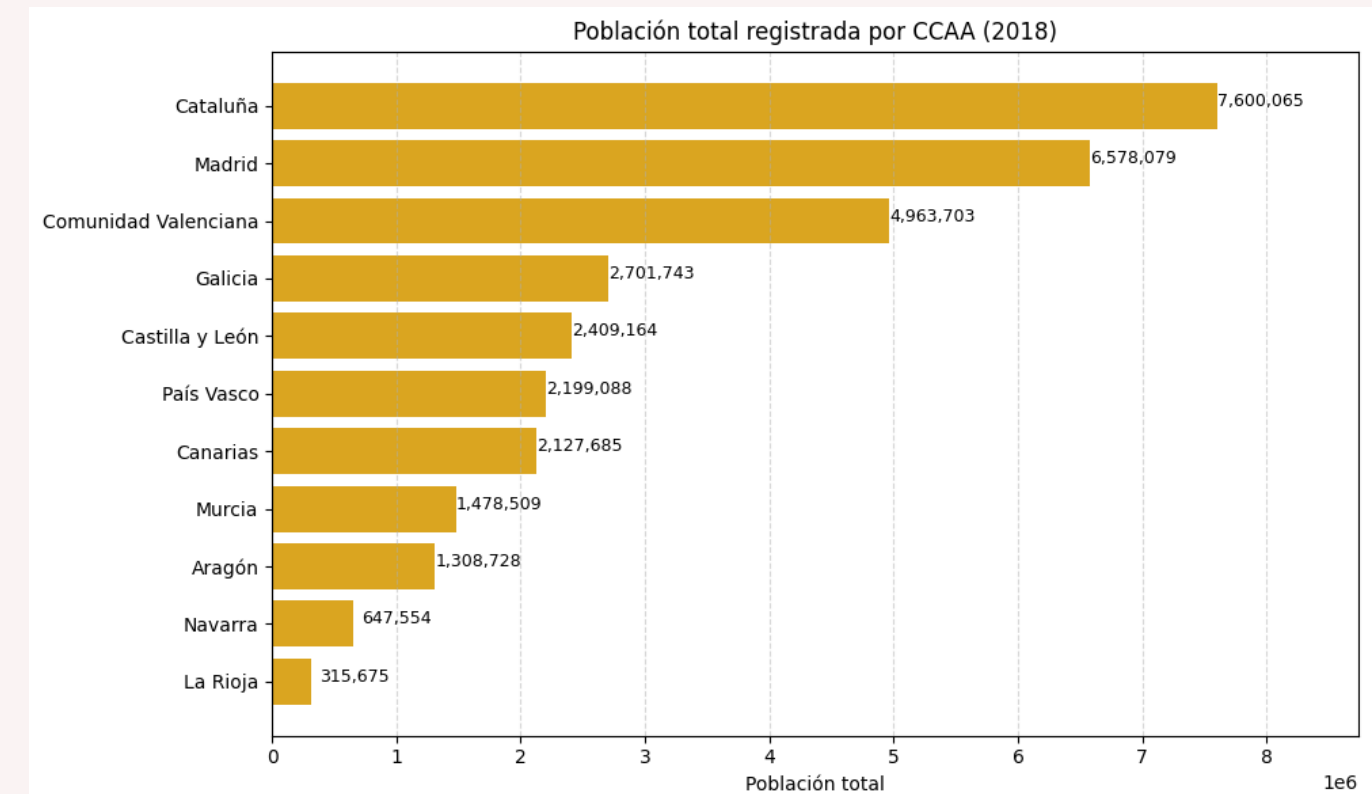
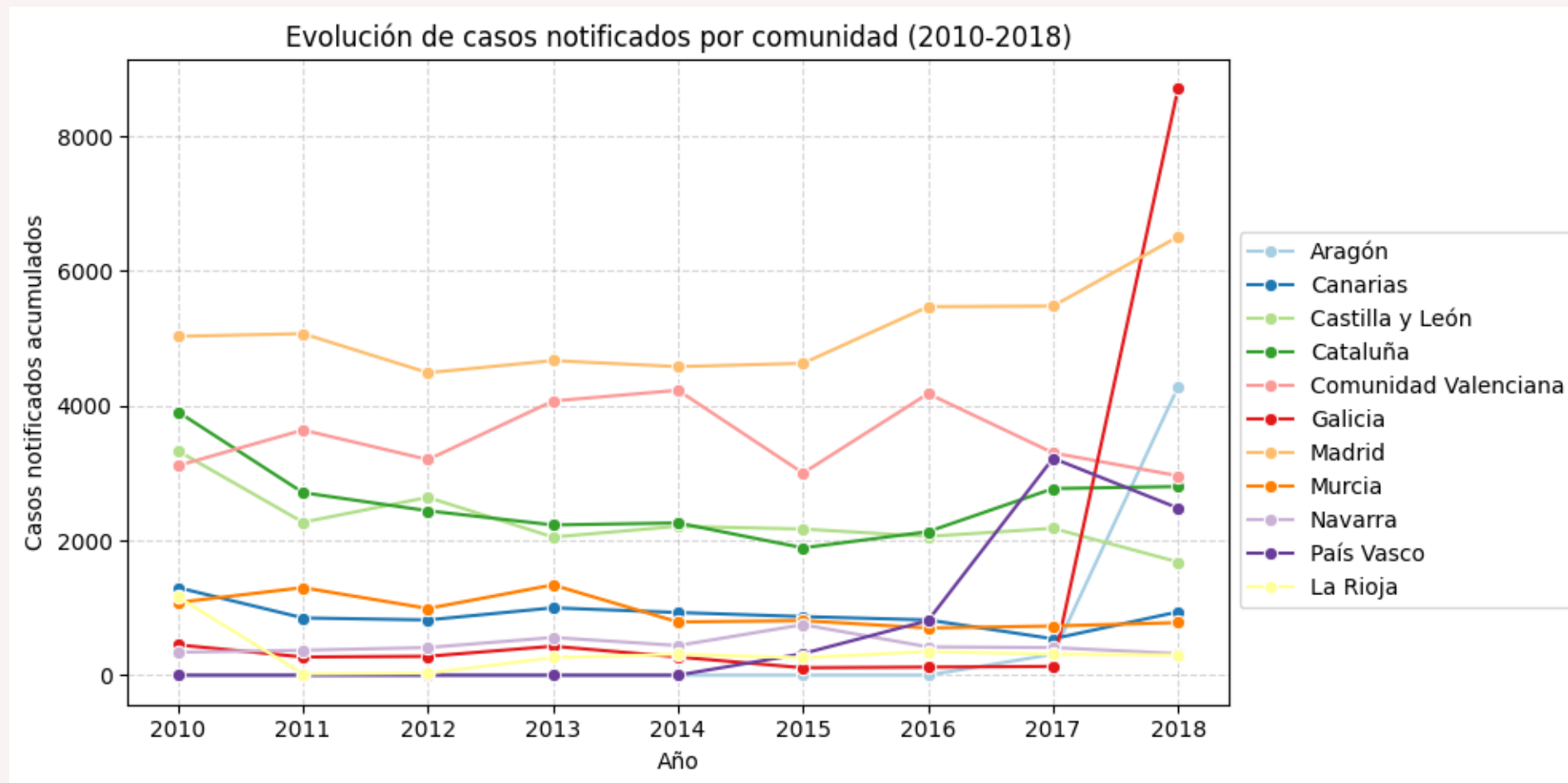
Gene Symbol	Full Name	Chromosome	Summary	NCBI Link
LMNA	lamin A/C	1	The protein encoded by this gene is part of the nuclear lam	<a href="https://www.ncbi.nlm.nih.gov/gene/">https://www.ncbi.nlm.nih.gov/gene/</a>
TP53	tumor protein p53	17	This gene encodes a tumor suppressor protein containing 1	<a href="https://www.ncbi.nlm.nih.gov/gene/">https://www.ncbi.nlm.nih.gov/gene/</a>



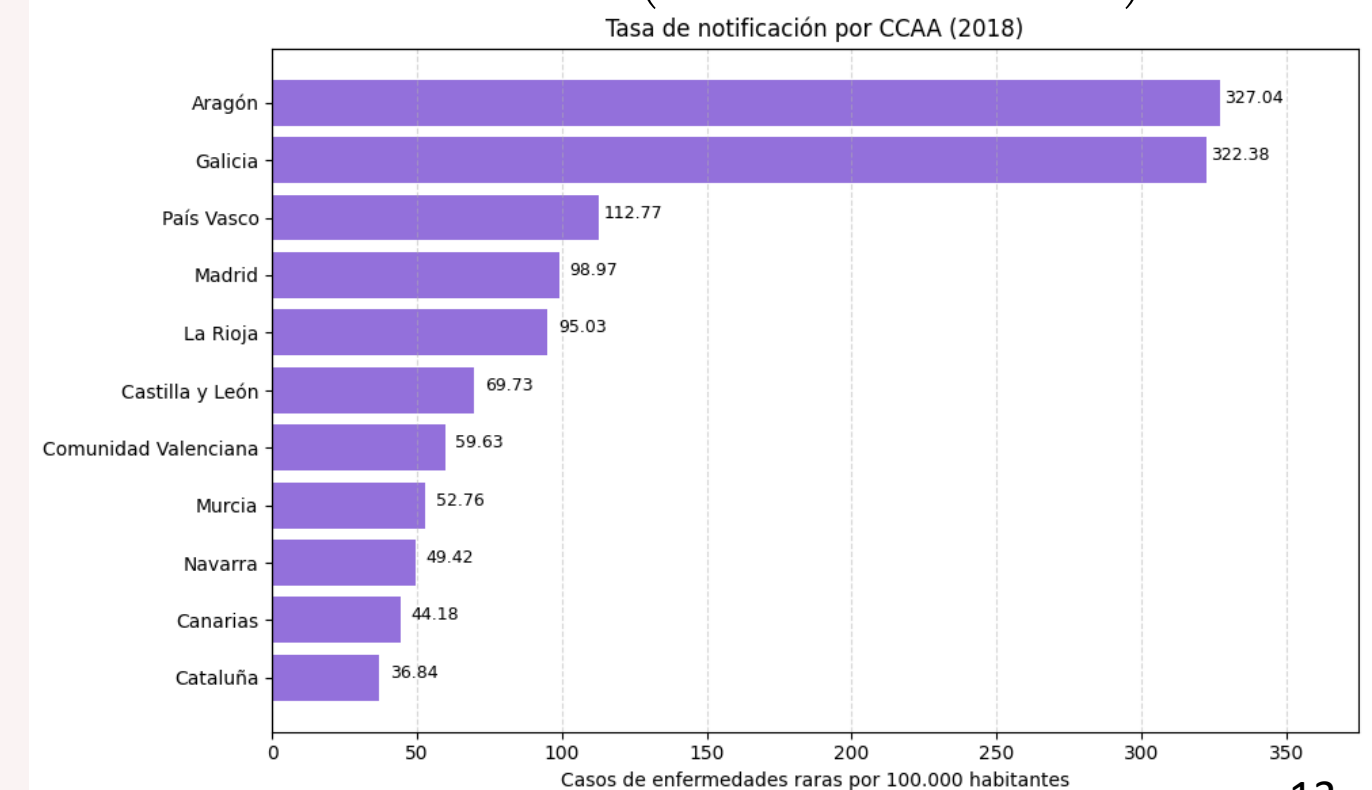
# Resultados

## - Reek -

### Incidencia de las ER en España (2010 - 2018)



$$\text{Tasa casos ER por 100.000 hab.} = \left( \frac{\text{Casos ER notificados en 2018}}{\text{Población total}} \right) \times 100.000$$

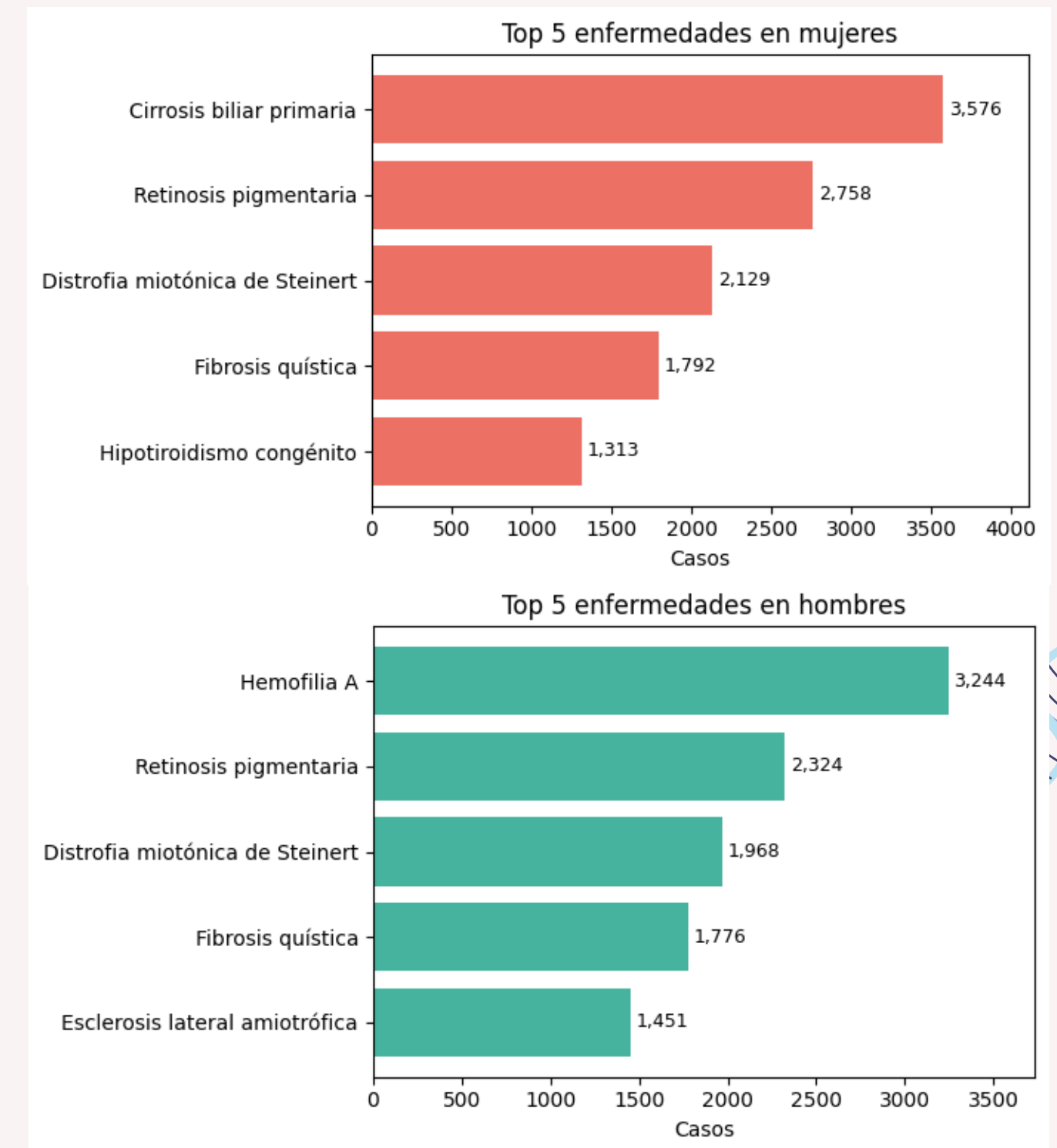
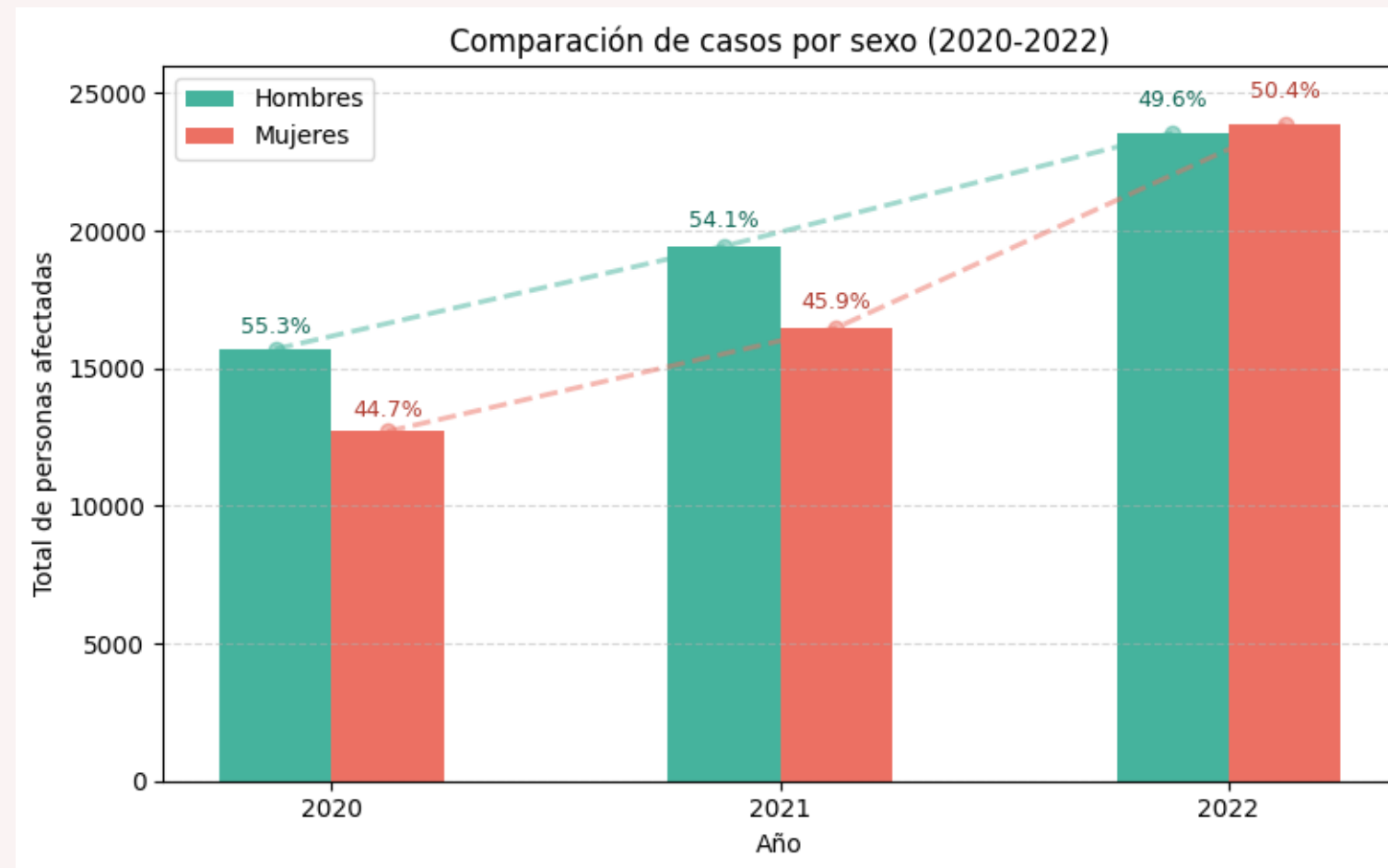


- ❑ Aumento de casos notificados en Galicia, Aragón y Madrid.
- ❑ Las comunidades con mayor tasa de notificación de ER por habitante son Aragón y Galicia, pero las más pobladas son Cataluña y Madrid.
- ❑ No hay relación directa entre cantidad de diagnósticos y población.

# Resultados

## - Reek -

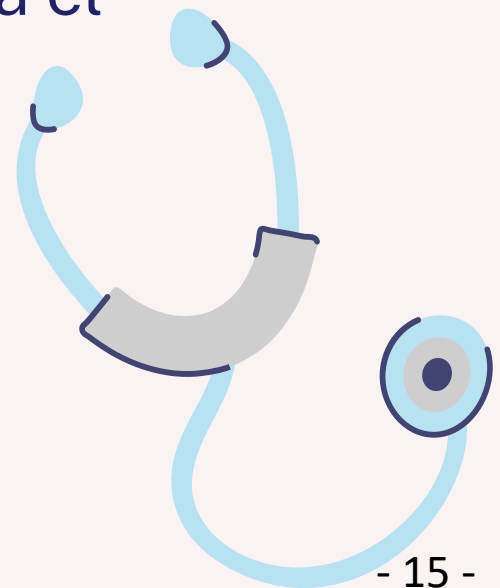
### Incidencia de las ER según género (2020 – 2022)

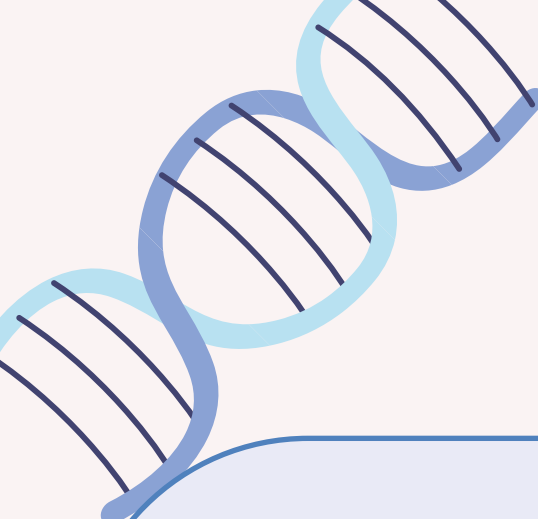


- ❑ Inicialmente los hombres presentaban una mayor incidencia (55%).
- ❑ En 2022 los diagnósticos en mujeres se incrementaron (50%).
- ❑ Tres de las cinco ER más notificadas presentan una incidencia similar.

# Conclusiones


- ❑ Se confirma el **crecimiento** sostenido de la **investigación** en el ámbito de las **Enfermedades Raras**.
- ❑ **No existe** una **correlación** significativa entre el número de **genes** y **fenotipos** asociados a las Enfermedades Raras de origen genético.
- ❑ En España, las comunidades autónomas con **mayor población** no concentran necesariamente **más diagnósticos**.
  - ❑ Posibles **diferencias** en la **organización sanitaria** a nivel autonómico. Desigualdad estructural.
- ❑ El **análisis por género** reveló un **aumento** significativo de diagnósticos en **mujeres**, hasta el 50 % en 2022.
  - ❑ Posibles mejoras en el acceso al diagnóstico o una reducción de sesgos clínicos.





# Trabajo futuro...

## Limitaciones del estudio realizado

- Fuentes de datos incompletas:
    - API de Orphadata aún en fase de implementación.
    - Falta de homogeneidad en los criterios de notificación al Registro Estatal de enfermedades Raras entre las diferentes comunidades.
  - Acceso limitado a los datos.
- 

- Clasificación de enfermedades por sistema afectado (Aprendizaje Supervisado).

Predecir a qué sistema del cuerpo afecta una ER a partir del estudio de características (prevalencia, edad de inicio, fenotipos, etc.).

- Analizar las interacciones entre un fármaco huérfano y la proteína de un gen.

Por ejemplo: interacción CBD y proteínas de los genes causantes de las EED (SYNGAP1, SLC6A1, etc.).

- Estudiar rutas metabólicas e interferencias con proteínas similares.

- Estudiar fármacos químicamente similares.
- 



**¡Muchas gracias!**

