Epidemiología oculta: exploración de las enfermedades raras en Europa y su impacto regional en España

o

Estudio descriptivo de las enfermedades raras en Europa y su impacto regional en España mediante datos públicos

## Resumen

## Introducción

¿Qué son las enfermedades raras? ¿Por qué son relevantes?

¿Cuáles son las dificultades del diagnóstico y tratamiento?

¿Cómo es su investigación? Hitos históricos.

¿Cuál es el papel del análisis de datos en este contexto?

\*\* Objetivo del trabajo: preguntas y/o hipótesis.

En Europa, se define enfermedad rara (ER) o poco frecuente como aquella que, con peligro de muerte o invalidez crónica, tiene una prevalencia menor de 5 casos por cada 10.000 habitantes1.

El Registro Estatal de Enfermedades Raras (ReeR) está formado por los registros autonómicos, encargados de transmitir los datos al primero.

Objetivo:

* Explorar y caracterizar la situación de las enfermedades raras en Europa y España a partir de datos públicos (Orphadata, ReeR, EMEA, PubMed), con el fin de identificar patrones epidemiológicos, genéticos y clínicos relevantes para su investigación y gestión sanitaria.

Preguntas e hipótesis:

* Las enfermedades raras con mayor número de fenotipos (HPO) tienden a tener mayor prevalencia, posiblemente porque son más fácilmente detectables o mejor documentadas.
* Las enfermedades raras del sistema nervioso son las más prevalentes, dado su impacto clínico y la frecuencia de estudio en investigaciones biomédicas.
* ¿Qué regiones de España presentan mayor carga de enfermedades raras según los informes ReeR?
* Las comunidades autónomas con mayor población no siempre concentran más diagnósticos, lo cual puede deberse a diferencias en el acceso al diagnóstico o registro.

## Metodología

Fuente de datos:

* API Orphadata
* European Medicines Agency (EMEA) Orphan designation
* Informes ReeR (Registro estatal enfermedades raras España) 2021 – 2024
* Publicaciones científicas de PubMed que contienen “rare disease”

Procesamiento de datos: limpieza + EDA + preprocesado

Herramientas utilizadas: Python (pandas, matplotlib/seaborn/plotly), Power BI, MySQL

Pipeline, filtros, etc.

**Fuente de datos**

**Herramientas utilizadas**

El desarrollo del proyecto se llevó a cabo utilizando Python como lenguaje de programación (versión 3.13.2), con Jupyter Notebook como framework principal para la ejecución interactiva del código y Visual Studio Code como entorno de programación.

Los análisis se realizaron empleando las siguientes bibliotecas:

* *pandas* para manipulación de datos tabulares
* *numpy* para operaciones numéricas
* *matplotlib* y *seaborn* para visualizaciones estáticas
* *plotly* para visualizaciones interactivas
* *scikit-learn* para análisis exploratorio y preprocesamiento
* *requests* y *xml.etree.ElementTree* para acceso y procesamiento de datos obtenidos por API

Todo el desarrollo se realizó en un entorno local, ejecutado sobre sistema operativo Windows 64 bits.

## Resultados

Resultados obtenidos, sin interpretación.

Tablas y gráficos: prevalencia, localización geográfica, edad, etc.

Distribución de enfermedades por región, clasificación según sistema afectado, según gen implicado, etc.

Evaluación según enfoque: epimediológicas/genéticas.

**Datos de Orphadata**

Para obtener datos de las enfermedades raras de Orphanet, se accedió a la API y se consiguió guardar nueve DataFrames en formato .csv.

Se estudió cuidadosamente el contenido de cada uno de estos DataFrames, se eliminaron las columnas innecesarias y se reorganizaron las otras, dando lugar a la siguiente base de datos conteniendo las columnas indicadas en la Figura 1

A computer screen shot of a computer

AI-generated content may be incorrect.

Figura 1. Diagrama relacional de la base de datos obtenida de Orphadata.

**Datos del registro estatal de enfermedades raras (ReeR)**

Se analizaron los informes del ReeR para obtener el total de casos notificados por comunidad autónoma entre 2010 y 2018. Se intentó estimar la evolución anual de nuevos diagnósticos mediante el cálculo de diferencias entre datos acumulados. Sin embargo, en varios casos se detectaron valores negativos, lo que sugiere modificaciones retrospectivas en los registros, más que una variación real en la incidencia. Por este motivo, se optó por no presentar este análisis y, en su lugar, mostrar la evolución de los casos notificados (Figura 2). En 2018 se observó un aumento considerable de casos en Galicia, Aragón y Madrid, que fueron las comunidades con mayor número de notificaciones durante ese año.Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Figura 2. Evolución del número de casos notificados por CCAA en el periodo 2010 - 2018

**Las comunidades autónomas con mayor población no siempre concentran más diagnósticos...**

Uno de los objetivos de este estudio fue analizar si las comunidades autónomas (CCAA) con mayor población concentran un mayor número de diagnósticos de enfermedades raras. Para ello, se utilizaron los datos de casos notificados por CCAA en 2018, extraídos del Registro Estatal de Enfermedades Raras (ReeR), y se contrastaron con la población total por CCAA en el mismo año, según datos del Instituto Nacional de Estadística (INE).

Con el fin de permitir una comparación proporcional entre regiones con diferente tamaño poblacional, se calculó una tasa estandarizada de notificación por cada 100.000 habitantes:

Los resultados se representaron mediante un gráfico de barras horizontales (Figura 3 y Figura 4).

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Figura 3. Población total por comunidad autónoma en el año 2018, registro extraído del INE

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Figura 4. Tasa de casos de enfermedades raras notificados por comunidad autónoma en 2018

El análisis mostró que comunidades con menor población, como Aragón y Galicia, presentaron tasas más altas de casos notificados. En contraste, Cataluña, una de las comunidades más pobladas, presentó la tasa más baja. Estos resultados sugieren que el volumen poblacional no se correlaciona necesariamente con una mayor proporción de diagnósticos notificados.

## Discusión

Análisis de los resultados.

Buscar patrones, tendencias y su significado.

Analizar si existe información suficiente para sacar conclusiones. Limitaciones del Dataset.

Analizar las implicaciones que tienen estos resultados para políticas de salud y para la investigación en este ámbito.

**Las comunidades autónomas con mayor población no siempre concentran más diagnósticos...**

Los resultados indican que la carga poblacional no se asocia directamente con una mayor proporción de diagnósticos notificados de enfermedades raras. De hecho, algunas comunidades con menor población (como Galicia y Aragón) presentan tasas más altas de notificación, lo que podría deberse a factores como una mayor eficiencia en los sistemas regionales de salud, mejores mecanismos de reporte o una mayor concienciación en profesionales sanitarios.

Por otro lado, la baja tasa observada en Cataluña podría reflejar subregistro, diferencias en criterios diagnósticos o barreras en la notificación. Estos hallazgos invitan a profundizar en las diferencias organizativas entre regiones y explorar qué factores pueden estar influyendo en las variaciones observadas, más allá del tamaño poblacional.

## Conclusión

Resumir los puntos clave de la investigación.

¿Qué aporta este trabajo?

Posibles mejoras y futuras líneas de investigación.

## Referencias