**Más que raras: exploración de las Enfermedades Raras en Europa y su impacto regional en España**

## Resumen

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Maecenas id imperdiet ante. Donec cursus nec urna eget sodales. Vestibulum massa felis, luctus a tellus non, tristique accumsan sem. Nam et eros vitae nibh vulputate elementum. Quisque nec diam sit amet urna sodales rutrum sed nec ligula. Quisque dictum nibh eu turpis fringilla, in dignissim diam ornare. Praesent aliquet, odio vel ullamcorper rhoncus, quam sem feugiat tellus, sed auctor ante augue eget sapien. Duis ut nisl ac lorem fermentum luctus quis id neque. Nam et arcu metus. Praesent porttitor, turpis sed scelerisque luctus, nisl purus rhoncus augue, a egestas odio libero ut urna. Vestibulum lacinia ex vel purus tincidunt, consequat facilisis ligula laoreet. Nulla porttitor, quam a sodales egestas, libero erat eleifend enim, sit amet imperdiet neque magna eu elit. In blandit, lorem at ultrices suscipit, erat orci feugiat nulla, at imperdiet erat neque eget ipsum. Pellentesque commodo, erat sit amet placerat gravida, tortor lacus finibus sapien, aliquet accumsan nisl quam a mauris.

## Introducción

Las enfermedades raras (ER) o enfermedades huérfanas, son aquellas que afectan a un número reducido de personas en relación con la población general. En Europa, se define como rara toda patología que afecta a menos de 5 por cada 10.000 habitantes (Orphadata, 2024; Posada et al., 2008). Aunque individualmente poco frecuentes, se estima que existen entre 6 – 10 mil ER, que en conjunto impactan a más de 400 millones de personas en todo el mundo (Mary Wang et al., s. f.; Venugopal et al., 2024), cifra comparable al impacto global al de las enfermedades comunes más prevalentes.

Más del 80 % de las ER son de origen genético y muchas se detectan en la infancia, lo que complica su identificación y diagnóstico clínico (Mary Wang et al., s. f.) A pesar de los avances, el diagnóstico de estas enfermedades sigue siendo muy complicado; en España, más del 56 % recibe el diagnóstico con un retraso superior al año, y en algunos casos, la espera supera los 10 años (Benito-Lozano et al., 2022). Este retraso conlleva consecuencias clínicas, psicológicas y sociales para los pacientes y sus familias.

La investigación en ER ha estado históricamente limitada por el escaso número de casos, la falta de modelos preclínicos y la baja rentabilidad percibida por la industria farmacéutica. Sin embargo, marcos regulatorios como la Orphan Drug Act (EE. UU., 1983) y la Regulación Europea de Medicamentos Huérfanos (EC 141/2000) han incentivado el desarrollo de fármacos huérfanos y aumentado la visibilidad científica del campo (Auvin et al., 2018).

En este contexto, el análisis de datos juega un papel cada vez más relevante. La incorporación de tecnologías de secuenciación masiva e inteligencia artificial permite acelerar la identificación de variantes genéticas, optimizar diagnósticos y desarrollar tratamientos personalizados (Banerjee et al., 2023; Wojtara et al., 2023). Estas herramientas son importantes dada la variabilidad clínica y la escasez de muestras.

Este estudio analiza la situación de las enfermedades raras en Europa y España utilizando datos públicos, con el objetivo de identificar patrones que contribuyan a su investigación y gestión sanitaria. Se examina si las enfermedades con más fenotipos están mejor caracterizadas, qué tipos presentan mayor diversidad genética y qué regiones españolas concentran más casos según el ReeR, especialmente en relación con el tamaño de su población.

## Metodología

**Fuente de datos**

Se utilizaron tres fuentes principales de información para el análisis. En primer lugar, se accedió a la API de PubMed para extraer publicaciones científicas que contienen en el título las palabras *“rare disease”*.

En segundo lugar, se utilizó la API de Orphadata para obtener el registro completo de enfermedades raras. Esta fuente proporciona información detallada sobre cada enfermedad, incluyendo su nombre, el identificador único (*ORPHAcode*), los genes asociados, los fenotipos relacionados, los distintos tipos de prevalencia y la edad de aparición.

Por último, se recopilaron datos del Registro Estatal de Enfermedades Raras (ReeR), que se construye a partir de los registros autonómicos, encargados de enviar los datos al Ministerio de Sanidad. Estos datos se presentan en informes periódicos en formato .pdf, publicados por el propio Ministerio. Se revisaron los informes correspondientes al periodo 2010–2022, y se extrajo información sobre los casos notificados en España. Además, los datos del año 2022 se contrastaron con las cifras de población total publicadas por el Instituto Nacional de Estadística (INE), con el objetivo de obtener una visión más precisa de la incidencia a nivel nacional.

**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 utilizando las bibliotecas *pandas* para la manipulación de datos tabulares, *numpy* para operaciones numéricas, *matplotlib* y *seaborn* para visualizaciones estáticas, y *requests* junto con *xml.etree.ElementTree* para el acceso y procesamiento de datos obtenidos a través de una API.

**Limpieza de datos**

Los datos de enfermedades raras se obtuvieron a través de la API de Orphanet, generando nueve archivos .csv. Se revisaron y limpiaron los DataFrames, eliminando columnas innecesarias, reorganizando campos y descartando valores nulos, excepto en el cálculo de la prevalencia final, que se filtrará más adelante por tipo de prevalencia. En el caso de los informes del ReeR, se cargaron las tablas, se adaptaron los formatos y se eliminaron los totales para evitar errores en el análisis.

## Resultados

**Datos de PubMed**

Tras el análisis de publicaciones científicas registradas en PubMed, se pudo observar que la primera vez que se habló de enfermedades raras fue en el año 1906, pero no fue hasta la década del 2010 que se vio un aumento significativo del número de investigaciones publicadas sobre enfermedades raras (Figura 1).

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Figura 1. Evolución de publicaciones científicas conteniendo "rare disease" en el título (datos extraídos de PubMed)

**Datos de Orphadata**

Tras analizar los datos recopilados de Orphadata, se obtuvo un total de XXXX ER distribuidas en 35 clases. La mayoría de las enfermedades entran en más de una clase, como se puede ver en la Figura 2, aunque se puede observar que las más abundantes son las genéticas.

Gráfico

El contenido generado por IA puede ser incorrecto.

Figura 2. Top 10 clases de ER más abundantes (izquierda) y distribución del número de clases al que pertenecen las ER (derecha)

**Datos del Registro Estatal de Enfermedades Raras (ReeR)**

Se analizaron los informes del ReeR para obtener el total de casos notificados por comunidad autónoma (CCAA) entre 2010 y 2018 (Grupo de trabajo del Registro Estatal de Enfermedades Raras, 2021) y se estudió la evolución de los casos notificados en este periodo de tiempo (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.

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Figura 3. Evolución del número de casos notificados por CCAA en el periodo 2010 - 2018

Uno de los objetivos de este estudio fue analizar si las comunidades 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, 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). 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.

Gráfico

El contenido generado por IA puede ser incorrecto.

Figura 4. Población total por CCAA registrada en 2018 según datos del INE (izquierda) y tasa de notificación de pacientes con enfermedades raras por cada 100.000 habitantes notificados según CCAA

A partir de los informes del Ministerio de Sanidad, se analizaron los casos notificados según el género para determinar si uno de los dos se ve más afectado por estas enfermedades. La Figura 4 muestra la comparación entre hombres y mujeres afectados entre 2020 y 2022(Grupo de trabajo del Registro Estatal de Enfermedades Raras, 2022, 2023, 2024). Aunque al inicio los hombres presentaban una mayor incidencia, en 2022 los diagnósticos en mujeres aumentaron significativamente, llegando a representar casi la mitad del total.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Figura 5. Comparación de diagnósticos por género en los años 2020 - 2022

También se analizó qué enfermedades afectan más a cada género, como se muestra en la Figura 5. Se observó que tres de las cinco enfermedades raras más notificadas presentan una incidencia similar en hombres y mujeres. Sin embargo, la Hemofilia A y la Esclerosis Lateral Amiotrófica afectan en mayor medida a los hombres, mientras que la Cirrosis Biliar Primaria y el Hipotiroidismo Congénito son más frecuentes en mujeres.

Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Figura 6. Comparativa de enfermedades raras más comunes por sexo (casos vivos en 2022)

## 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

Auvin, S., Irwin, J., Abi-Aad, P., & Battersby, A. (2018). The Problem of Rarity: Estimation of Prevalence in Rare Disease. *Value in Health*, *21*(5), 501-507. https://doi.org/10.1016/j.jval.2018.03.002

Banerjee, J., Taroni, J. N., Allaway, R. J., Prasad, D. V., Guinney, J., & Greene, C. (2023). Machine learning in rare disease. *Nature Methods*, *20*(6), 803-814. https://doi.org/10.1038/s41592-023-01886-z

Benito-Lozano, J., López-Villalba, B., Arias-Merino, G., Posada de la Paz, M., & Alonso-Ferreira, V. (2022). Diagnostic delay in rare diseases: data from the Spanish rare diseases patient registry. *Orphanet Journal of Rare Diseases*, *17*(1). https://doi.org/10.1186/s13023-022-02530-3

Grupo de trabajo del Registro Estatal de Enfermedades Raras. (2022). *Informe ReeR 2022: Situación de las Enfermedades Raras en España*. https://www.sanidad.gob.es/areas/alertasEmergenciasSanitarias/vigilancia/docs/Informe\_Epidemiologico\_Anual\_2022\_ReeR\_ACCESIBLE.pdf

Grupo de trabajo del Registro Estatal de Enfermedades Raras. (2023). *Informe ReeR 2023: Situación de las Enfermedades Raras en España*. https://www.sanidad.gob.es/areas/alertasEmergenciasSanitarias/vigilancia/docs/InformeEpidemiologicoAnual\_2023\_ACCESIBLE.pdf

Grupo de trabajo del Registro Estatal de Enfermedades Raras. (2024). *Informe ReeR 2024: Situación de las Enfermedades Raras en España*. https://www.sanidad.gob.es/areas/alertasEmergenciasSanitarias/vigilancia/docs/InformeEpidemiologicoAnual\_2024\_ACCESIBLE.pdf

Grupo de trabajo del Registro Estatal de Enfermedades Raras. (2021). *Informe ReeR 2021: Situación de las Enfermedades Raras en 2010‐2018*. https://www.sanidad.gob.es/areas/alertasEmergenciasSanitarias/vigilancia/docs/Informe\_reer\_reducido\_2021.pdf

Mary Wang, C., Wainstock, D., Dawkins, H., Hon Yin Chung, B., & Tsz Wai Chu atwchu, A. (s. f.). *Rare disease emerging as a global public health priority*.

Orphadata. (2024, julio). *Free access products description* (Orphadata, Ed.). https://www.orphadata.com/docs/OrphadataFreeAccessProductsDescription.pdf

Posada, M., Martín-Arribas, C., Ramírez, A., Villaverde, A., & Abaitua, I. (2008). Enfermedades raras: Concepto, epidemiología y situación actual en España. *Anales del Sistema Sanitario de Navarra*. https://scielo.isciii.es/scielo.php?script=sci\_arttext&pid=S1137-66272008000400002

Venugopal, N., Naik, G., Jayanna, K., Mohapatra, A., Sasinowski, F. J., Kartha, R. V., & Rajasimha, H. K. (2024). Review of methods for estimating the prevalence of rare diseases. *Rare Disease and Orphan Drugs Journal*, *3*(1). https://doi.org/10.20517/rdodj.2023.39

Wojtara, M., Rana, E., Rahman, T., Khanna, P., & Singh, H. (2023). Artificial intelligence in rare disease diagnosis and treatment. En *Clinical and Translational Science* (Vol. 16, Número 11, pp. 2106-2111). John Wiley and Sons Inc. https://doi.org/10.1111/cts.13619