

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

## Resumen

Este estudio analiza las enfermedades raras (ER) en Europa y España a partir de datos públicos provenientes de PubMed, Orphadata y el Registro Estatal de Enfermedades Raras (ReeR). Se investigó la relación entre el número de genes y fenotipos asociados a cada enfermedad genética, sin encontrarse correlación significativa. Además, se identificaron los genes y enfermedades más frecuentemente asociados. A nivel regional, se observó que comunidades con menor población presentan tasas más altas de notificación de ER, lo que sugiere posibles diferencias en la eficiencia del sistema de salud o en los mecanismos de reporte. También se analizaron las diferencias por género en el diagnóstico. Los resultados evidencian desigualdades regionales en la detección y registro de ER, lo que subraya la necesidad de fortalecer los sistemas de vigilancia y diagnóstico en España.

## Introducción

Las enfermedades raras afectan a menos de 5 personas por cada 10.000 habitantes (Orphadata, 2024; Posada et al., 2008). Se estima que existen entre 6 – 10 mil ER, que en conjunto impactan a más de 400 millones de personas en todo el mundo (Chung et al., 2022; 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, 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).

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 accedió a la API de PubMed para extraer publicaciones científicas que contienen en el título las palabras “rare disease”.

También se utilizó la API de Orphadata para obtener el registro completo de enfermedades raras. Esta fuente contiene 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 ReeR, formado a partir de los registros autonómicos, quienes los datos al Ministerio de Sanidad. Estos datos se presentan en informes (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), para tener una visión más precisa de la incidencia a nivel nacional.

## Limpieza de datos

El desarrollo del proyecto se llevó a cabo utilizando *Python* como lenguaje de programación (versión 3.13.2). 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.

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

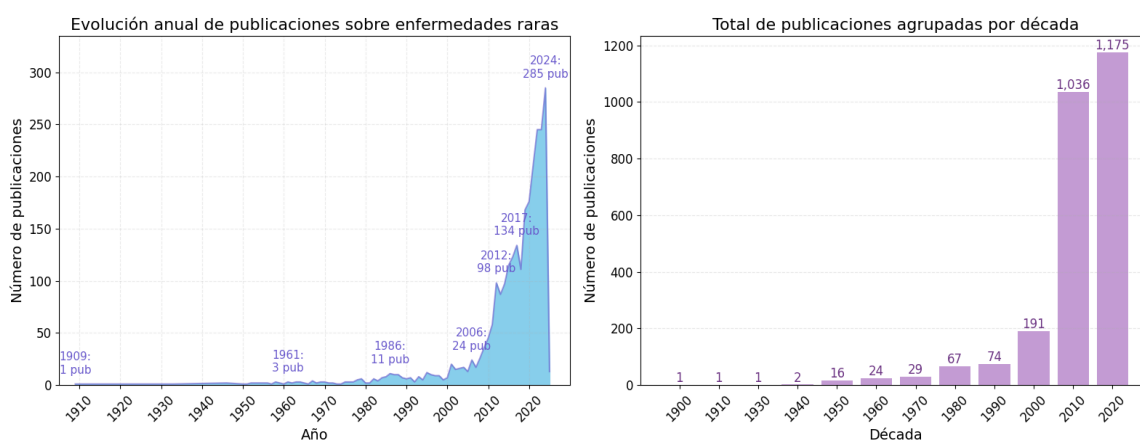


Figura 1. Evolución de publicaciones científicas conteniendo "rare disease" en el título (datos extraídos de PubMed)

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; **Error! No se encuentra el origen de la referencia.**).

### Datos de Orphadata

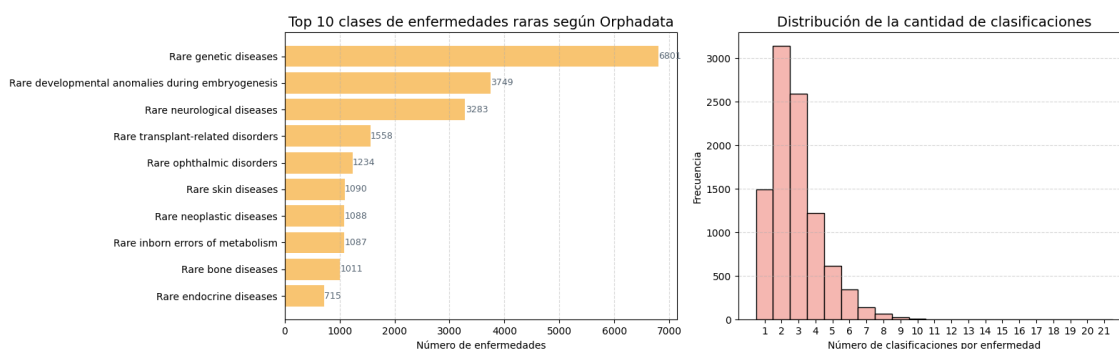


Figura 2. Top 10 clases de ER más abundantes (izquierda) y distribución del número de clases al que pertenecen las ER (derecha)

Tras analizar los datos recopilados de Orphadata, se obtuvo un total de 9673 ER distribuidas en 35 clases. La mayoría de las enfermedades entran en más de una clase, como se puede ver en la Figura 2, y se puede observar que las más abundantes son las genéticas.

Se analizaron únicamente enfermedades genéticas, evaluando la cantidad de fenotipos y genes asociados a cada una. Se utilizó un gráfico *jointplot* (Figura 3) para mostrar la distribución conjunta del número de genes y fenotipos por enfermedad. El gráfico de dispersión central permitió explorar si existe relación entre ambas variables, pero no se observó una tendencia clara. La correlación de Spearman entre el número de genes y fenotipos fue de 0.0392, indicando una relación prácticamente nula.

También se identificó la enfermedad con mayor cantidad de genes asociados: la Discapacidad intelectual no síndromica (ORPHAcode: 528084), con 108 genes distintos. Por otro lado, los genes LMNA y TP53 son los que están implicados en más enfermedades, con 21 asociaciones cada uno.

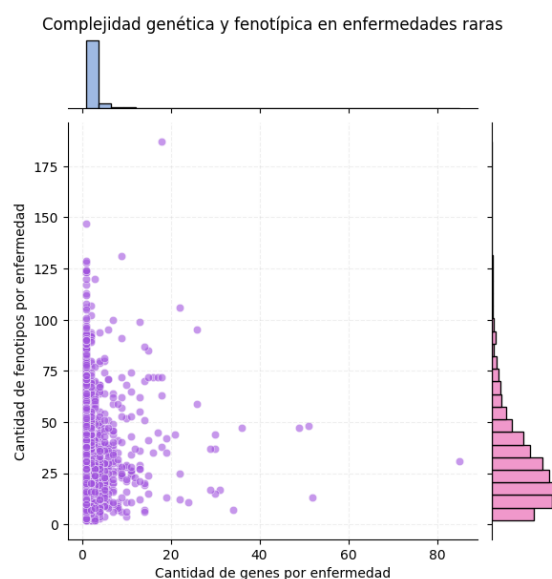


Figura 3. Relación entre el número de genes y de fenotipos asociados a ER genéticas

### 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 ). En 2018 se observó un aumento considerable de casos en Galicia, Aragón y Madrid, que fueron las comunidades con mayor número de notificaciones.

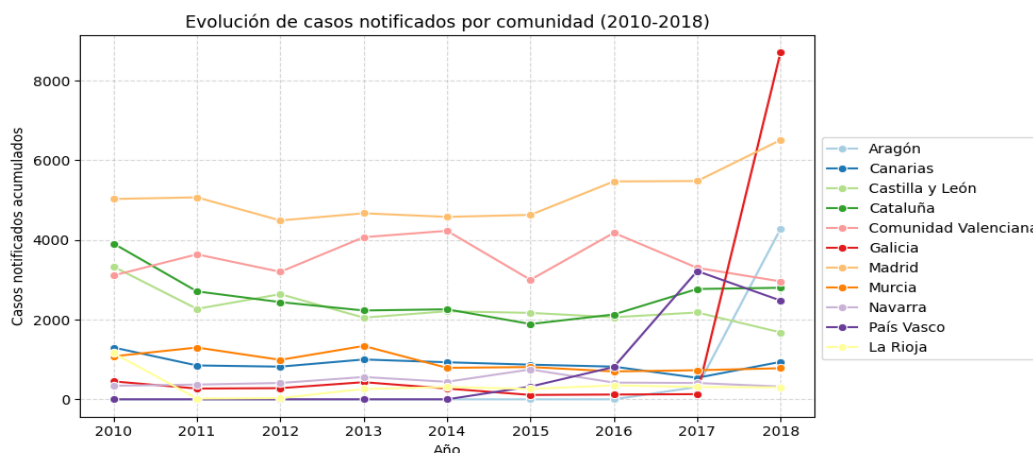


Figura 4. Evolución del número de casos notificados por CCAA en el periodo 2010 - 2018

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:

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

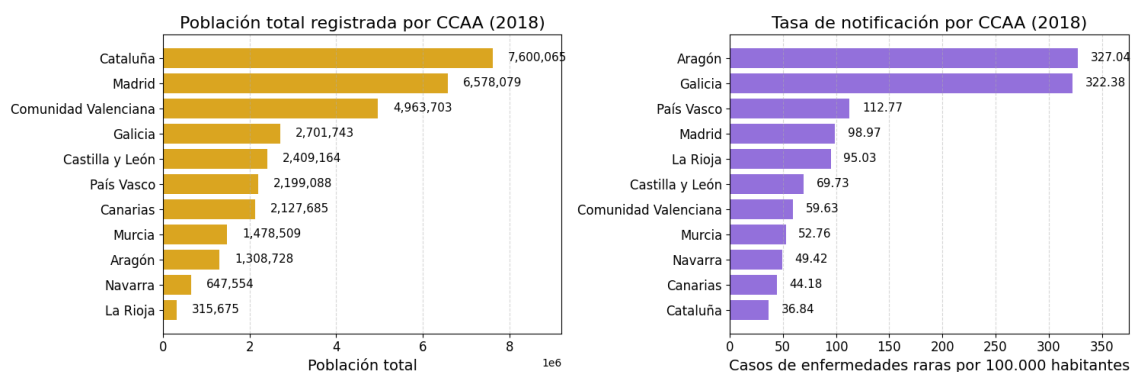


Figura 5. Población total por CCAA registrada en 2018 según datos del INE (izquierda) y tasa de notificación de pacientes con enfermedades raras por cada 100.000 habitantes notificados según CCAA

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

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 6 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.

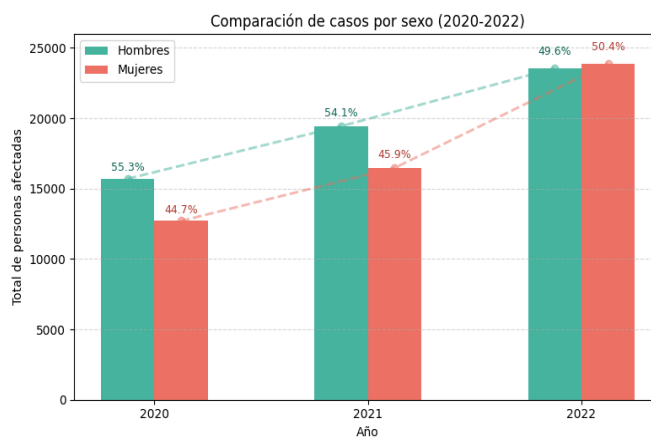


Figura 4. Comparación de diagnósticos por género en los años 2020 - 2022

También se analizó qué enfermedades afectan más a cada género, como se muestra en la Figura 7. 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

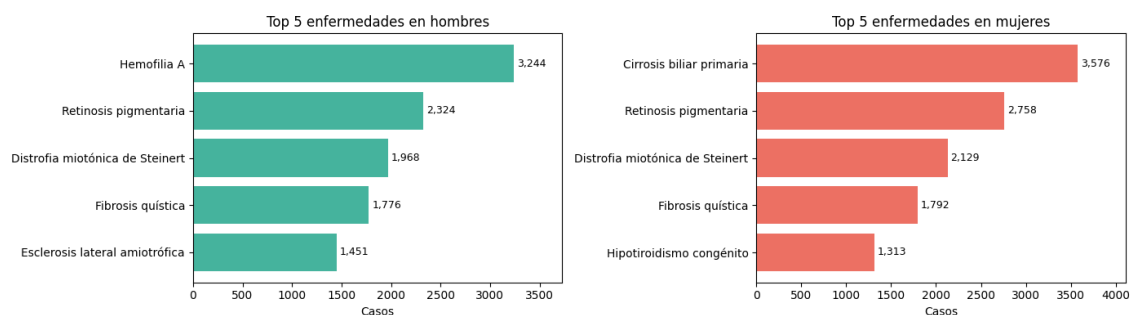


Figura 5. Comparativa de enfermedades raras más comunes por sexo (casos vivos en 2022)

mujeres.

## Discusión

Los resultados muestran que el interés científico en las enfermedades raras ha crecido de forma sostenida, especialmente a partir de la década de 2010. Este aumento en las publicaciones

registradas en PubMed refleja una mayor visibilidad del campo, impulsada por avances tecnológicos, marcos regulatorios específicos y una creciente demanda social.

En cuanto al análisis gen-fenotipo, se observó una amplia variabilidad en el número de genes y fenotipos por enfermedad, con una correlación prácticamente nula entre ambas variables, lo que indica que una alta complejidad genética no necesariamente se traduce en una alta diversidad fenotípica, y viceversa.

Los datos del ReeR revelan que las comunidades autónomas con mayor población no concentran necesariamente más diagnósticos. De hecho, regiones como Galicia y Aragón (con menor densidad poblacional) reportaron tasas proporcionalmente más altas de notificación. Esto podría deberse a una mejor organización de sus sistemas de salud, mayor cobertura diagnóstica o simplemente a un registro más riguroso de los casos. Por el contrario, la baja tasa de notificación en Cataluña sugiere posibles limitaciones en los sistemas de reporte, diferencias en criterios diagnósticos o incluso falta de concienciación. Estas variaciones apuntan a desigualdades estructurales entre comunidades que deben abordarse si se busca una estrategia sanitaria equitativa en el ámbito de las ER.

Finalmente, el análisis por género mostró que, aunque históricamente los hombres presentaban una mayor incidencia, en los últimos años el número de diagnósticos en mujeres ha aumentado, alcanzando casi la mitad del total en 2022. Esta tendencia podría reflejar una mejora en el acceso al diagnóstico o cambios en los sesgos clínicos, pero requiere más investigación.

## Conclusión

Este trabajo aporta una visión integrada de las enfermedades raras desde distintas dimensiones: evolución científica, complejidad genético-clínica y distribución regional en España. Se confirma el crecimiento sostenido de la investigación en el ámbito y se revela que no existe una correlación significativa entre el número de genes y fenotipos por enfermedad. Además, se evidencian desigualdades territoriales importantes en la notificación de casos, que no se explican únicamente por el tamaño poblacional.

Estos hallazgos pueden servir como punto de partida para mejorar la gestión sanitaria de las enfermedades raras, promoviendo sistemas de vigilancia más eficientes, mejor formación de profesionales y políticas que reduzcan las brechas regionales.

## 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. (2021-2024). Informe ReeR 2022: Situación de las Enfermedades Raras en España.
- Chung, C. C. Y., Hong Kong Genome Project, Chu, A. T. W., & Chung, B. H. Y. (2022). Rare disease emerging as a global public health priority. *Frontiers in public health*, 10, 1028545.
- 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](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>