# Scaling Rare Disease detection across multiple pathologies and sites using Generative AI and OMOP CDM

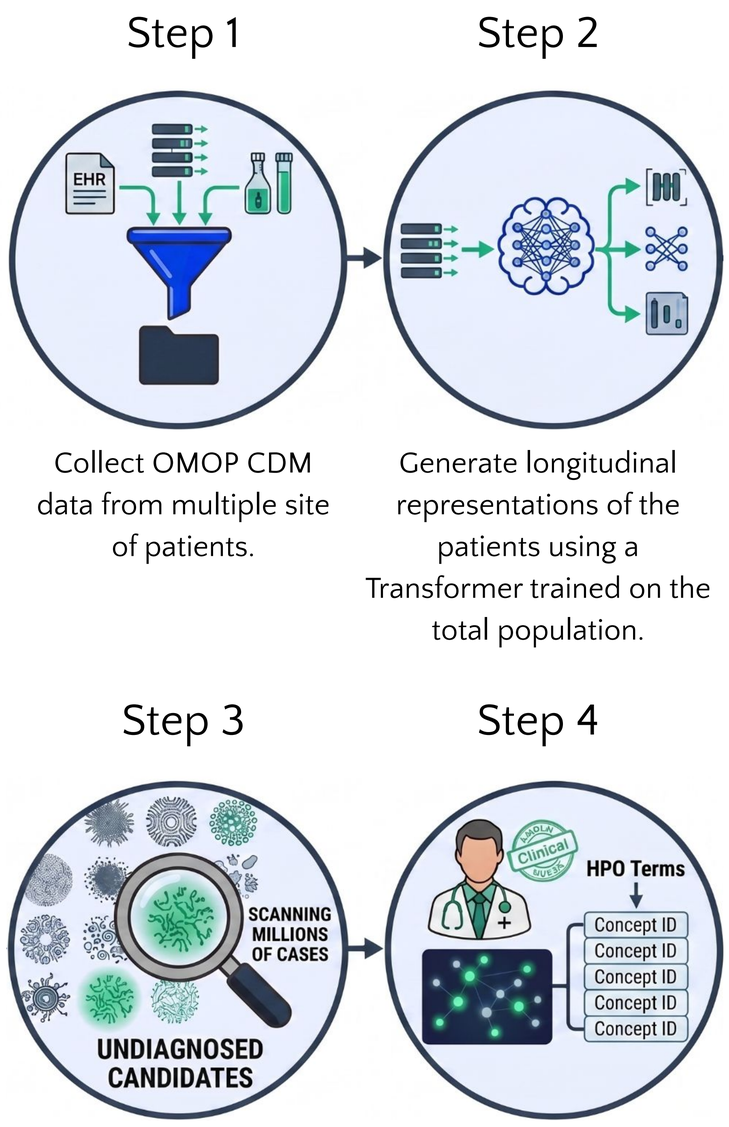
## *A multicentric patient finder study for Rare Disease on Nephrology*


**Background:**

Hereditary Kidney Diseases (HKD) present with scattered, non-specific symptoms that evolve slowly, resulting in an average diagnostic delay of 5–7 years. Previous work[1] showed that AI could identify undiagnosed Alport Syndrome patients, but the static approach treated symptoms as an unordered list and failed to generalize to other hospitals and diseases with complex temporal patterns. We developed a Longitudinal Patient Similarity (LPS) framework, testing whether time and context-aware patient trajectories could detect four distinct pathologies on patients with Chronic Kidney Disease (CKD) across multiple hospitals.


## Methods:

1. **Data architecture & interoperability:** We leveraged EHR data (clinical notes, laboratory results, and pharmacology) already mapped to the OMOP CDM to build the longitudinal patient sequences and ensure cross-institutional normalization and data quality across 3 sites. Human Phenotype Ontology (HPO) terms were used by the clinicians to define the phenotypes of interest and these were mapped to OMOP CDM concepts.
2. **Longitudinal patient sequence modeling:** We employed transformers to generate a mathematical representation of the patient longitudinal data that captured the temporal dependencies and pathophysiological trajectories (e.g., recognizing prodromal acroparesthesia preceding clinical nephropathy in Fabry disease). References to the target pathologies were removed from the sequences.
3. **Executing the patient finder:** We then executed the search based on seed cohorts (e.g. n=10–22) of patients with a confirmed diagnosis to establish a phenotype query, enabling the efficient screening of a federated population dataset (N=1.2M, 72,611 with CKD).
4. **Explainable AI (XAI):** The search algorithm used aCompressed Latent Reasoning (CLaRa)[2] to decode the model's latent representation, mapping the predictive clinical signals of the matched patients back to the standardized HPO terms initially used by the clinicians, helping the clinicians with interpretation during the validation.
5. **Validation of the found patients with the Rare Disease:** The LPS framework was retrospectively validated across four distinct hereditary conditions: Alport Syndrome, Fabry Disease, Primary Hyperoxaluria, and Tuberous Sclerosis Complex (TSC).




**Results and Conclusions:**

- Achieved >84% sensitivity across all cohorts in target population of 72,611 patients with CKD.
- Median Number Needed to Screen (NNS) of 30 patients to identify one true case.
- Currently deployed in routine practice to shorten diagnostic delays on one pathology

| Disease | TP | Seed | S / TPR | NNS |
|---|---|---|---|---|
| **Fabry Disease** | 53 | 13 | 81.6% | 31.99 |
| **Alport Syndrome** | 408 | 88 | 85.7% | 27 |
| **Primary Hyperoxaluria** | 14 | 3 | 84.6% | 140 |
| **TSC** | 151 | 19 | 85.8% | 29.1 |

Table 1. Performance Metrics across validation. Seed represents the initial group of confirmed patients used to initialize the model, which then identifies True Positives (TP). Efficiency is measured by Sensitivity (S / TPR), the ability to correctly flag at-risk patients. The Number Needed to Screen (NNS) indicates how many patients must be evaluated to identify one true case


[1] Elizabeth R. Viera et al., Leveraging AI for early detection of chronic kidney and hereditary kidney diseases: a focus on Alport syndrome, Nephrology Dialysis Transplantation, Volume 40, October 2025, gfaf116.1061, https://doi.org/10.1093/ndt/gfaf116.1061
[2] Jie He et al. CLaRa: Bridging Retrieval and Generation with Continuous Latent Reasoning https://doi.org/10.48550/arXiv.2511.18659



**Gabriel Maeztu** gabriel.maeztu@iomed.health
**Mónica Arrúe** monica.arrue@iomed.health
**Roser Torra** rtorra@fundacio-puigvert.es
**Josep Cordón** josep.cordon@iomed.health
**Elizabeth R. Viera** erviera@fundacio-puigvert.es
**Leonor Fayos** lfayos@fundacio-puigvert.es