

Towards a Pan-European Generative AI for Population Health

Tuomo Hartonen and **Jonas Burian** and **Tatiana Cajuso** and **Matteo Ferro** and **Andrea Ganna** *Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland.*
{tuomo.hartonen, andrea.ganna}@helsinki.fi
Taavi Tillmann and **Nikita Umov** *University of Tartu, Tartu, Estonia.*
Francesca Ieva and **Andrea Corbetta** and **Emanuele Di Angelantonio** *Human Technopole, Milan, Italy.*
Oliver Stegle *European Molecular Biology Laboratory, Heidelberg, Germany.*
Adrien Coulet and **Emmanuel Bacry** and **Tim Vlaar** and **Valentine Facque** and **Mario Jendrossek** *Health Data Hub, Paris, France.*
Pascaline Aucouturier and **Clément Henin** and **Sandrine Katsahian** *Greater Paris University Hospitals (APHP), Paris, France.*
Anton Geerinck and **Bert Vaes** and **Martina Fonseca** *Sciensano, Brussels, Belgium.*
Troels Siggaard and **Lisa Cantwell** and **Soren Brunak** *University of Copenhagen, Copenhagen, Denmark.*
Pero Ivanko *Croatian Institute for Public Health, Zagreb, Croatia.*
Hannah Brooke and **Tove Fall** *Uppsala University, Uppsala, Sweden.*
Moritz Gerstung *German Cancer Research Center, Heidelberg, Germany.*

Abstract

Generative Artificial Intelligence (GenAI) has the potential to transform population health research by enabling digital twins and disease modelling across countries. Realizing this vision in Europe requires harmonized use of health data within the European Health Data Space (EHDS). We present first steps towards a pan-European GenAI framework. Using aggregated data from 84M individuals across nine countries, we evaluated cross-country comparability of diagnosis and medication coding, revealing good concordance across Europe but marked divergences with U.S. data, underscoring the need for EU-specific models. Pooling data can enable rare disease studies that would be otherwise infeasible. Delphi, a foundation model trained on UK health trajectories performed robustly in Finland, though with imperfect transferability. These findings establish the feasibility of pan-European GenAI and outline a roadmap for developing foundation models aligned with EHDS objectives. Our ongoing work includes integration of more diverse data modalities, and improving GenAI model transferability.

1 Introduction

Biomedical research is being reshaped by computational advances and growing availability of multimodal health data, including electronic health records (EHRs) and nationwide registers. The EHDS aims to provide a secure and

interoperable framework for health data exchange, empowering citizens and enabling secondary use. Expanding access to such data can support trustworthy AI systems that improve healthcare across Europe (Ganna et al. [2024]). By learning from large, heterogeneous datasets, GenAI models can simulate individual health trajectories, support digital twin technologies, and enable precision medicine at population scale. Yet, most existing models are trained on EHRs from a single healthcare provider, often in the United States (US), limiting their generalizability (Waxler et al. [2025]). We present initial results towards building a pan-European GenAI model for population health: 1) We evaluate consistency in disease and prescription patterns across nine European national and regional cohorts (~ 84 M individuals), 2) assess the transferability of a UK-trained foundation model for health trajectories (Delphi, Shmatko et al. [2025]), to Finnish nationwide data, and 3) expand Delphi to model more complex health trajectories, including information on both diagnoses and medications.

2 Data Resources

Finnish Nationwide Dataset: Our comprehensive dataset of ~ 7.1 M individuals (similar to Viippola et al. [2023]) comprises hospital visits, medication purchases, and laboratory measurements for a total of ~ 532 million tokens.

Pan-European Cohorts: We considered nine datasets: national registries (Finland, Denmark, Estonia, Sweden, Croatia, France, Belgium), a regional cohort (Lombardy, Italy), and a biobank-based resource (All of Us, USA). We identified individuals aged 19–69 years and alive on January 1, 2010, and followed them until December 31, 2019. We recorded the number of individuals with first occurrences of diagnoses (3-digit ICD-10) and first prescriptions (5-digit ATC). The combined dataset totalled 84M individuals (Figure 1A).

3 Results

3.1 Cross-Country Comparability

Correlations between country pairs were consistently stronger for medication prescription than for diagnosis rates (Figure 1B). For medication prescriptions, the highest agreement was observed between Finland and Sweden, reflecting similarities in healthcare systems and prescribing practices. By contrast, the weakest correlations for diagnoses were observed between the US and European countries. Age-specific cardiovascular disease incidence increased steadily with age, with strong agreement across the European datasets (Figure 1C). Incidence rates were systematically lower in the US-cohort, likely reflecting differences in healthcare access and the limitations of insurance-based data.

Finally, the pan-European scale of the data enabled analyses of rare diseases that would be infeasible in individual countries. For example, by aggregating cases of Cushing’s syndrome across cohorts, we achieved sufficient sample sizes to support predictive modeling and trajectory analysis (Figure 1D).

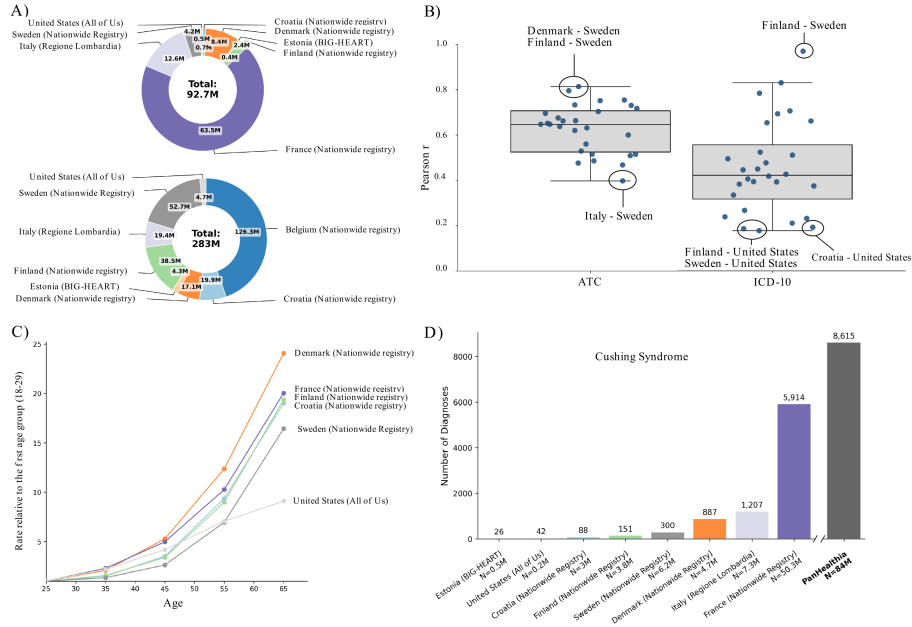


Figure 1: Preliminary results from 9 countries across 84 million individuals. A) Number of diagnoses (upper panel) and medication prescriptions/purchases (lower panel) by country. B) Pearson correlations of occurrence rates for medication purchases/prescriptions (left, ATC) and diagnoses (right, ICD). C) Age-specific cardiovascular disease rates across countries. D) Number of Cushing's syndrome diagnoses by country and in the combined cohort (PanHealthia).

3.2 Delphi GenAI Model Transferability to Finnish data

Delphi is a GenAI model designed specifically for longitudinal health trajectories and trained on 402,799 trajectories from the UK Biobank (UKB). In Delphi, clinical diagnoses are tokenized into sequences, enabling learning temporal dependencies across 1,257 different ICD-10 codes. Delphi has demonstrated strong performance in cross-disease prediction and in generating realistic patient-level disease trajectories, when evaluated on 100,639 individuals from the UKB, and 1,930,000 individuals from Danish registries (Shmatko et al. [2025]).

We evaluated Delphi transferability on Finnish nationwide data. Delphi achieved robust overall predictive accuracy, though with slightly lower performance compared to a Delphi-FIN model trained on Finnish health trajectories (Figure 2A). We then expanded Delphi-FIN vocabulary by including medication purchases using 5-digit ATC-codes. This increased the Delphi-FIN performance in future disease prediction across the ICD-chapters (Figure 2B).

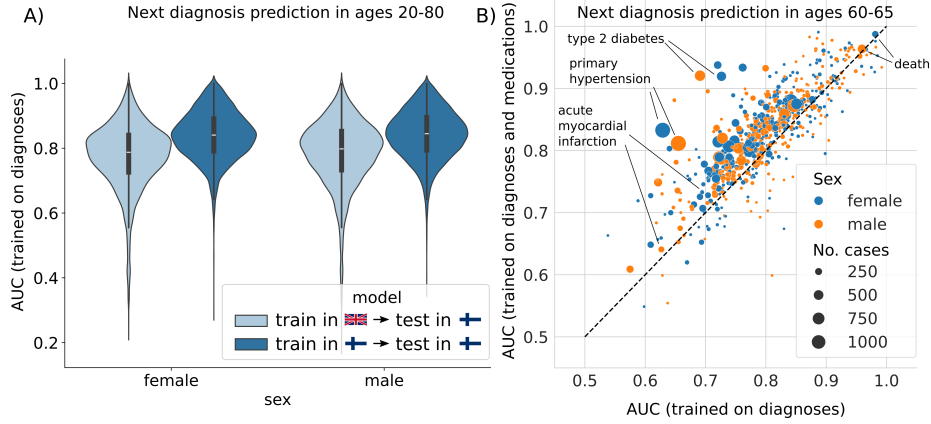


Figure 2: GenAI model performance evaluated on 1 million Finnish health trajectories. A) Distributions of next diagnosis prediction AUCs over 1,257 diseases. Light blue = Delphi trained on 402,799 UKB health trajectories, dark blue = Delphi-FIN trained on 402,799 Finnish trajectories. B) Next diagnosis prediction AUC in 60-65-yo individuals. Delphi-FIN was trained on 5 million Finnish trajectories (x-axis: diagnoses, y-axis: diagnoses and medications).

4 Conclusions and Future Directions

European healthcare systems routinely collect valuable health information. Modern AI-based generative models such as Delphi hold vast potential in modelling future trajectories both on individual and population levels. For individuals, these models could offer more personalized risk estimation, on a population level they could be used, for example, to better allocate healthcare resources. Healthcare GenAI models such as CoMET (Waxler et al. [2025]) developed on massive US-datasets are not only usually proprietary, but as we show, the US health data is differently distributed from European data. Even though there are less differences between European countries, transferability of models across many European countries will still need to be demonstrated. Here, we show that the Delphi model trained using UK data performs well in a nationwide Finnish cohort, although still not comparably to a model trained on Finnish data. These findings demonstrate the potential for GenAI models to generalize across European countries while also highlighting current limitations in cross-country differences in diagnostic coding practices and healthcare structures.

To enhance the transferability of GenAI models, we are implementing a method that maps healthcare system codes into a shared embedding space, leveraging semantic similarity and ontological relationships between codes, and building on earlier work (Kirchler et al. [2025]). We will further integrate diverse data modalities to improve model performance and systematically evaluate transferability across more than ten EU countries.

References

- A. Ganna, A. Carracedo, C. F. Christiansen, E. Di Angelantonio, P. A. Dykstra, A. M. Dzhambov, R. Eils, S. Green, K. L. Schneider, T. V. Varga, et al. The european health data space can be a boost for research beyond borders. *Nature Medicine*, 30(11):3053–3056, 2024.
- M. Kirchler, M. Ferro, V. Lorenzini, C. Lippert, A. Ganna, et al. Large language models improve transferability of electronic health record-based predictions across countries and coding systems. *medRxiv*, 2025.
- A. Shmatko, A. W. Jung, K. Gaurav, S. Brunak, L. H. Mortensen, E. Birney, T. Fitzgerald, and M. Gerstung. Learning the natural history of human disease with generative transformers. *Nature*, pages 1–9, 2025.
- E. Viippola, S. Kuitunen, R. S. Rodosthenous, A. Vabalas, T. Hartonen, P. Vartiainen, J. Demmler, A.-L. Vuorinen, A. Liu, A. S. Havulinna, et al. Data resource profile: Nationwide registry data for high-throughput epidemiology and machine learning (finregistry). *International Journal of Epidemiology*, 52(4):e195–e200, 2023.
- S. Waxler, P. Blazek, D. White, D. Sneider, K. Chung, M. Nagarathnam, P. Williams, H. Voeller, K. Wong, M. Swanhorst, et al. Generative medical event models improve with scale. *arXiv preprint arXiv:2508.12104*, 2025.