

## “Understanding Treatment Dynamics and Response from Population Electronic Health Records and Biobank Data”

The increased availability of health data has opened new possibilities to study treatments in everyday life. The thesis examines how medical treatments are implemented and effective in the real world by integrating population-scale electronic health records (EHRs) and biobank resources. It advances two tightly linked themes: treatment dynamics (initiation, persistence, adherence) and treatment response, arguing that real-world effectiveness depends as much on the continuity and implementation of care as on biological efficacy. To support credible inference from routine data, the work emphasises scalable, harmonised, and interpretable analyses that bridge heterogeneous health systems and national contexts.

Methodologically, the thesis leverages time-varying survival models for evolving, schedule-based exposures, Functional Data Analysis to represent long-horizon medication trajectories, and augments these with genetic epidemiology (genome-wide association studies and polygenic scores) when biobank data are available. Empirically, the thesis spans complementary data infrastructures and questions. Using Lombardy’s linked administrative EHRs, it quantifies the landscape and consequences of COVID-19 undervaccination with a dynamic exposure definition. In Finnish nationwide registers and linked biobanks (with replication in Estonia), it characterises medication adherence across common drug classes, separating persistence from adherence, and tests whether genetic predisposition adds explanatory or predictive value beyond social and clinical factors. Furthermore, it defines and investigates socioeconomic and health and determinants in more than one million individuals across Finland and Italy. Finally, with FinnGen’s laboratory and genetic data, it integrates short- and long-term LDL-cholesterol responses to statins with adherence, dosage, and pharmacogenomics, showing when genetics informs baseline risk and when treatment implementation dominates observed outcomes.

Across these applications, the thesis provides practical guidance for extracting and curating EHR information in Italy, reusable representations and tools for dynamic treatments, and cross-cohort comparisons that distinguish between shared regularities and context-specific features. The overarching contribution is a set of population-scale, reproducible approaches that turn routine health data into interpretable evidence on who starts, continues, and benefits from therapy; clarify the limited but targeted roles for genetics in predicting response in treated populations; and inform policies and clinical strategies to support sustained implementation, thereby moving from average efficacy toward dependable effectiveness in everyday practice.