A new regulatory-grade chronic disease cohort simulator: leveraging the large, deep and long CONSTANCEs cohort linked and matched to the SNDS in France

MSR7
ISPOR
Europe
2025

Authors: Marie Génin, MSc[1], Antoine Movschin, MSc[1], Sofiane Kab, PharmD, PhD[2], Billy Amzal, MBA, MPH, MSc, PhD[1]
Affiliation: [1]Quinten Health, Paris, France ; [2]INSERM, Villejuif, France

INTRODUCTION

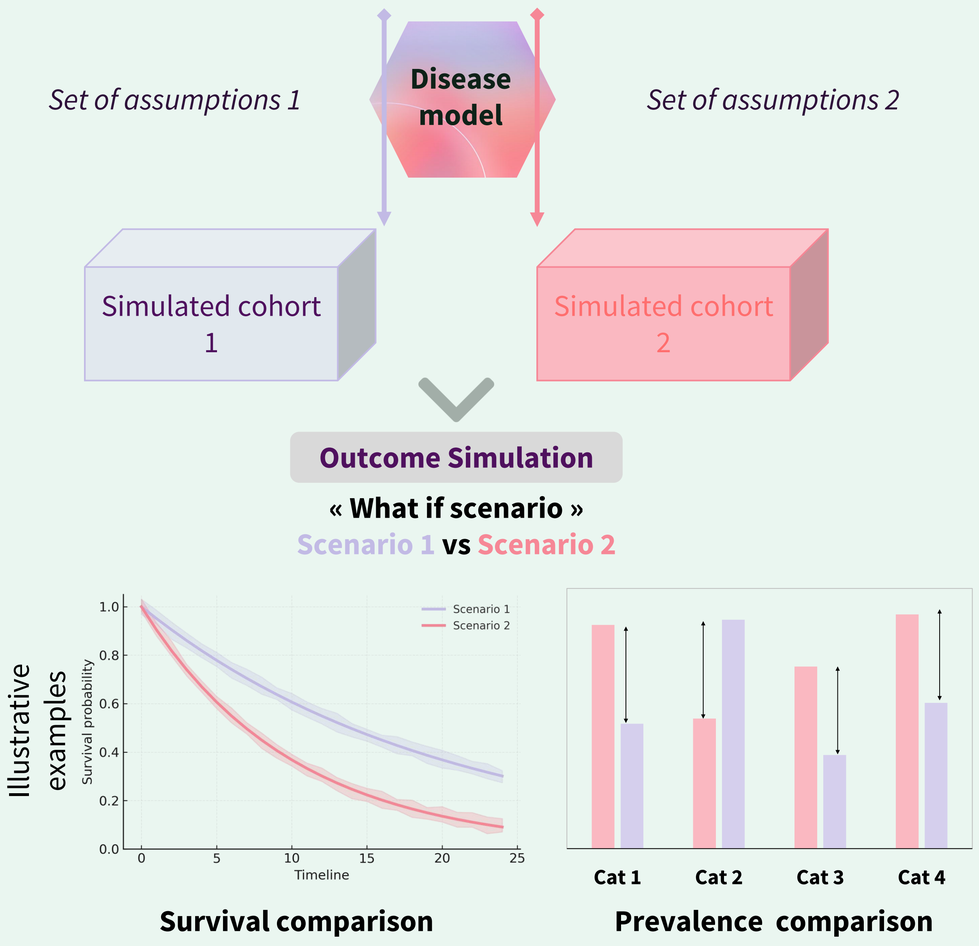
- RWD are increasingly recognized as essential to support decision-making in drug development, market access, and HTA.
- However, the use of such data faces persistent challenges, including data fragmentation, privacy constraints, and limited generalizability of existing cohorts.
- Virtual cohorts and synthetic populations have recently emerged as innovative solutions to overcome these challenges, enabling simulation of clinical and economic outcomes under realistic conditions while preserving patient privacy[1].
- Quinten Health, in collaboration with INSERM/CONSTANCES, is developing a regulatory-grade, disease-specific simulator built on the linked CONSTANCES–SNDS cohort[2].
- The simulator leverages advanced AI and modeling approaches to accelerate and de-risk evidence generation for regulatory and HTA purposes (illustrative use case of the framework is shown in Figure 1).

OBJECTIVE

Develop and validate a synthetic cohort simulator that:

- Identify patient profiles for clinical and access decisions.
- Characterize disease progression from longitudinal real-world data.
- Emulate SoC and treatment exposure.
- Simulate alternative clinical or population scenarios.
- Model costs and quality-of-life outcomes to support payer and policy evaluation.

Figure 1: Illustrative use case of the virtual cohort simulator



METHODS

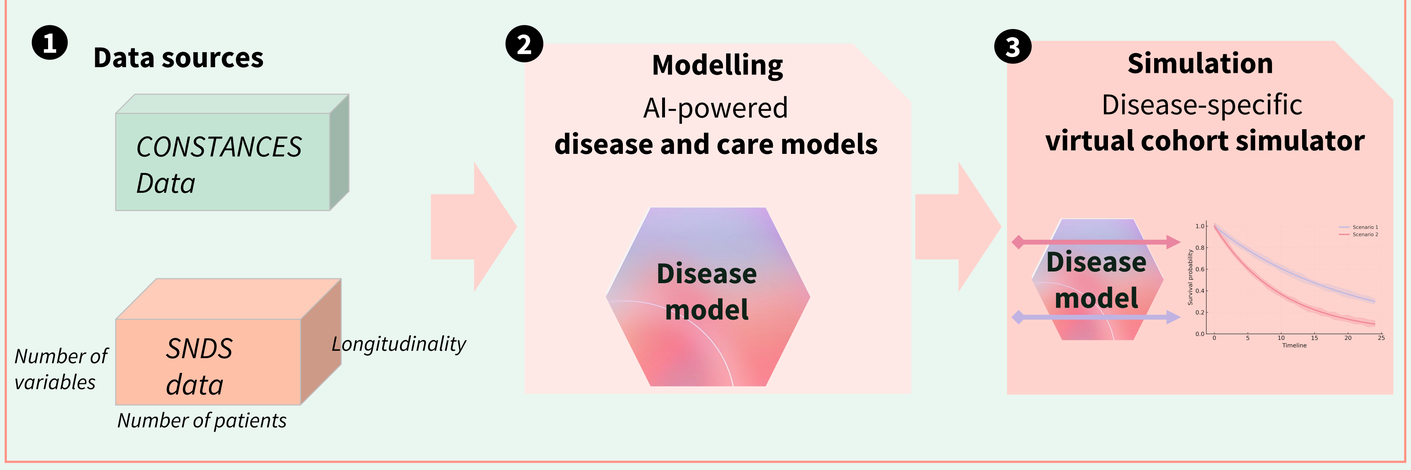
The simulator combines data from the CONSTANCES cohort (>220 000 adults) and the French National Health Insurance Database (SNDS), linking clinical, biological, socioeconomic, and healthcare information with over ten years of follow-up.

It is structured in three layers (Figure 2):

1. Data preparation – Integration and harmonization of CONSTANCES and SNDS data.
2. Modeling – Development of AI-powered disease and care models to describe patient trajectories and treatment response.
3. Simulation – Generation of disease-specific virtual cohorts for scenario analyses.

The AI-powered core (Figure 3) combines clustering, Bayesian/survival models, neural networks, and cost-effectiveness models to produce synthetic, privacy-preserving populations (via GANs) and HTA-relevant outcomes such as quality of life, resource use, and medical costs.

Figure 2: General concept of the cohort simulator



RESULTS

Feasibility and scope

- The CONSTANCES-SNDS cohort enables reconstruction of >10 years of longitudinal care data, covering clinical, biological, and administrative dimensions.
- Preliminary exploration confirms sufficient coverage and granularity to model disease trajectories in COPD and obesity.
- Core methodological modules – clustering and disease progression modelling – have been initiated and are under technical evaluation.

Early proof-of-concept work

- In COPD, inhaled therapy sequences were extracted and clustered using ML to identify patient trajectories.
- In obesity, initial disease-progression models were tested to simulate BMI evolution and comorbidity onset.
- These findings suggest the feasibility of transforming real-world data into structured patient profiles, supporting simulator calibration.

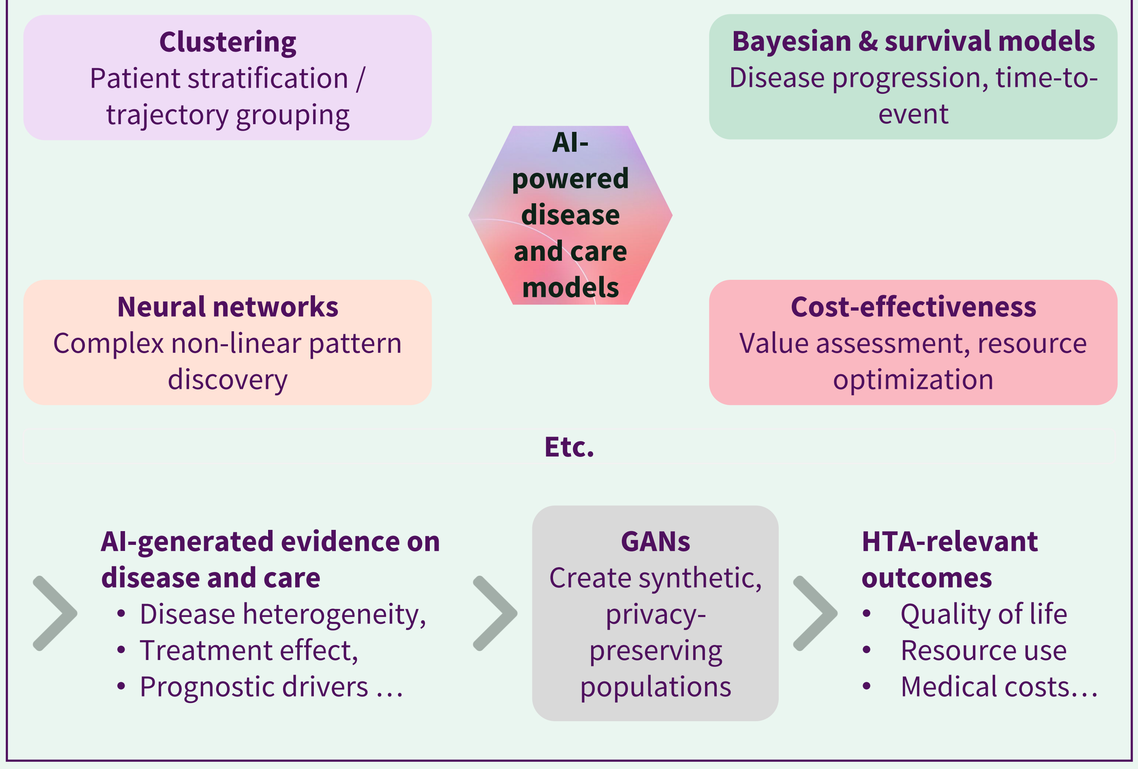
Ongoing developments

- Development of synthetic population generation (GANs) and disease-progression models (Bayesian and survival-based) is underway for COPD and obesity.
- Next steps include:
  - Validation of simulated outcomes against real-world data.
  - Calibration of economic and quality-of-life modules.
  - Integration of disease-specific models into a unified simulator.

Expected outcomes

A regulatory-grade virtual cohort simulator capable of reproducing real-world trajectories, testing alternative treatment scenarios, and supporting HTA and market-access evaluations.

Figure 3. Structure of the AI-powered modeling core



CONCLUSION

- The CONSTANCES-SNDS integration creates a unique, fit-for-purpose dataset with depth, breadth, and longitudinal follow-up, ideal for early decision-making in drug development, access, and evaluation.
- This regulatory-grade simulator aligns with HTA and payer expectations, offering credible tools to simulate real-world impacts, mitigate uncertainties, and guide strategic choices from target population definition to reimbursement planning.
- Future expansions will broaden applications to additional chronic diseases, fostering precision medicine that benefits patients, healthcare systems, and industry stakeholders through ethical, data-driven innovations.

REFERENCES:

1. Goncalves A, et al. Generation and evaluation of synthetic patient data. Nature Communications. 2020;11:2083.
2. Zins, M. et al. The French CONSTANCES Population-Based Cohort: Design, Inclusion and Follow-Up. https://doi.org/10.1007/s10654-015-0096-4

ABBREVIATIONS:
AI: Artificial Intelligence; COPD: Chronic Obstructive Pulmonary Disease; GANs: Generative Adversarial Networks HTA: Health technology assessment; RWD: Real-world data; SNDS: Système National des Données de Santé; SoC: Standard of Care.

CONFLICT OF INTEREST: NA
ACKNOWLEDGEMENT: We would like to thank Sonia Zebachi for her support in drafting and reviewing the submitted abstract for ISPOR EU 2025.

CONTACT INFO: m.genin@quinten-health.com

quinten health