Work Package #1: Health Sustainability
The health data platform

Tiziano Arduini[1], Vincenzo Atella [1], Federico Belotti[1 2], Andrea Piano Mortari [1 2 3]

[1]Department of Economics and Finance, University of Rome Tor Vergata

[2]CEIS Tor Vergata

[3]Ministry of Health, Department of Programming

October 31, 2023

Outline WP #1

1. WP 1 objectives
2. The data
3. The Epidemiological Atlas at Province level: Imputing health indicators in missing provinces
4. Projecting population health status and expenditure: The Future Italian Model
   - Model structure and goals
   - Data
   - Projection and Validation
5. The Health-lab dashboard within AMELIA: an example
6. Long term sustainabilty / Business opportunities
   - The possibility of generating a Spin-off
   - A "Real world" example of what the Spin-off can do

Outline

WP 1 objectives

- The work package objective is to analyse the connections between **health shocks**, **socio-economic factors** (income, wealth, education, household composition), **environmental variables**, and **policies** that can protect households.
- To achieve this objective, a **Health Lab** will be created, integrating and disseminating a collection of diverse geo-referenced health and socio-economic databases.
- These databases will then be used **to monitor the Italian population** health status and the utilization of healthcare services of Italian households.

WP 1 activities

List of activities:

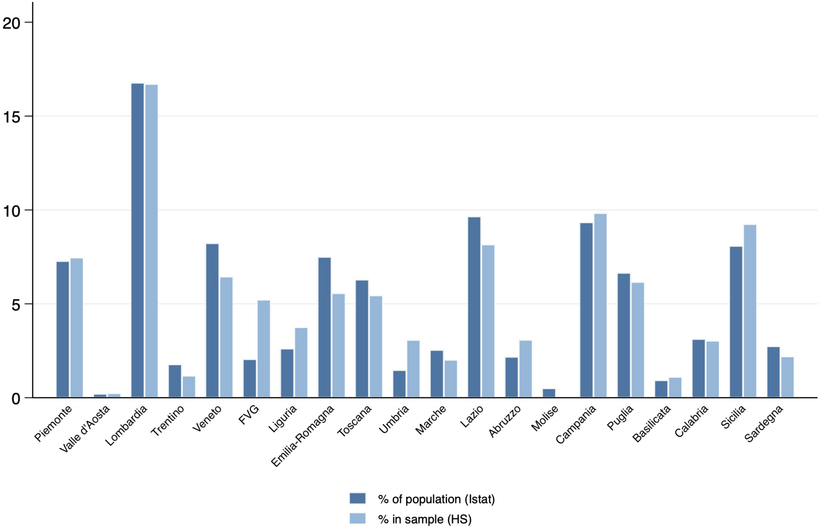
- production of **forecasts and scenarios** of population health status, needs and access to services at local level (province/Local Health Authority);
- measurement of **health inequality** by gender, age, income, region, citizenship, and other socio-economic variables;
- evaluation of the **effects of socio-economic and environmental factors**, as well as **welfare policies**, on household health status, wealth and labor supply;
- impact of health and other exogenous shocks on **children's** education, well-being, and cognitive as well non-cognitive outcomes.

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR
Finanziato dall'Unione europea
NextGenerationEU
GRINS

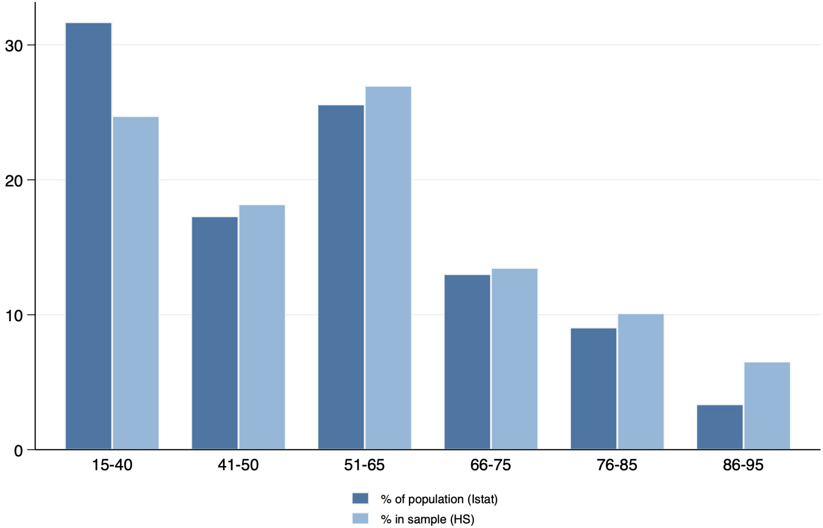
The Health Search/IQVIA Health LPD Longitudinal Patient Database (HS), an Italian general practice registry, which collects ECR data from patients aged 14+ registered with GPs:

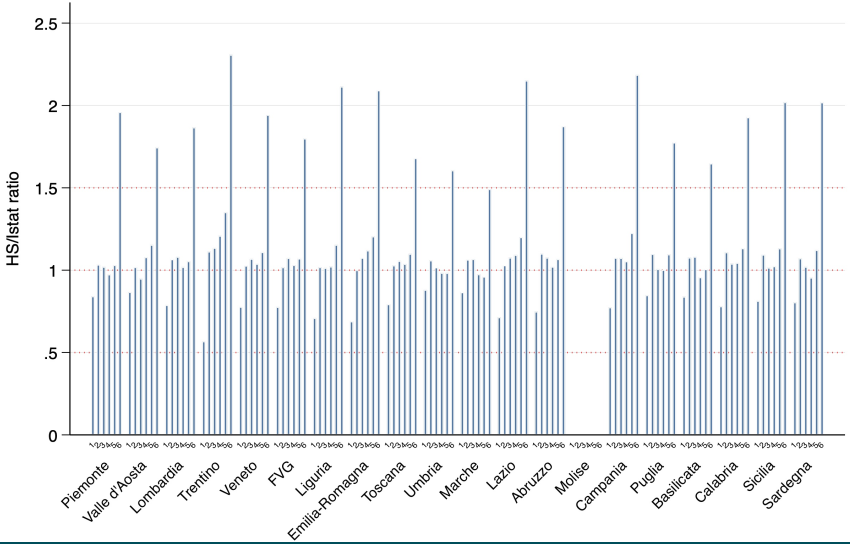
- 1250 GPs selected at regional level;
- More than 2 million patients (age and gender distributions aligned with ISTAT official statistics), of which about 90% followed for $\geq$ 10 years;
- Information on disease diagnoses (ICD-9 code);
- Information on drug prescriptions and diagnostic tests (AIC, ATC and National Tariff code);
- The database complies with European Union guidelines on the use of medical data for research;
- A key feature of the HS database is that it includes all patients registered in the GP lists, thus avoiding selection bias based on health status, which is a rather standard problem with claim data.

HS vs Istat (2021)



HS vs Istat (2021)




HS/Istat ratio
0
.5
1
1.5
2
2.5
Piemonte
Valle d'Aosta
Lombardia
Trentino
Veneto
FVG
Liguria
Emilia-Romagna
Toscana
Umbria
Marche
Lazio
Abruzzo
Molise
Campania
Puglia
Basilicata
Calabria
Sicilia
Sardegna

Outline

Imputing health indicators

Two issues for the imputation of health indicators at the province level:

- Spatial and serial correlation of health conditions (i.e., virus contagion and/or unobserved environmental exposures).

- Missing data issue is exacerbated when spatial interactions are at work.

For the imputation, we adopt a spatial Durbin autoregressive model with missing data in the outcome:

$$\boldsymbol{y} = \lambda_0 \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{W}\boldsymbol{X}\gamma_0 + \epsilon,$$

where,

- $\boldsymbol{y}$ is the health indicator, a n-vector at the province level for a specific **age group** and **gender**: missing in some provinces;
- $\boldsymbol{W}$ is the $n \times n$ normalized spatial (contiguity) matrix;
- $\boldsymbol{X}$ is a $n \times K$ matrix of predictors observed for all provinces.

Given that both predictors and the contiguity matrix are observed for all provinces, we exploit the reduced form of the model to perform two tasks simultaneously:

- predict the health indicators for the missing provinces;
- consistently estimate the coefficients (Wang and Lee, Econom J 2013).

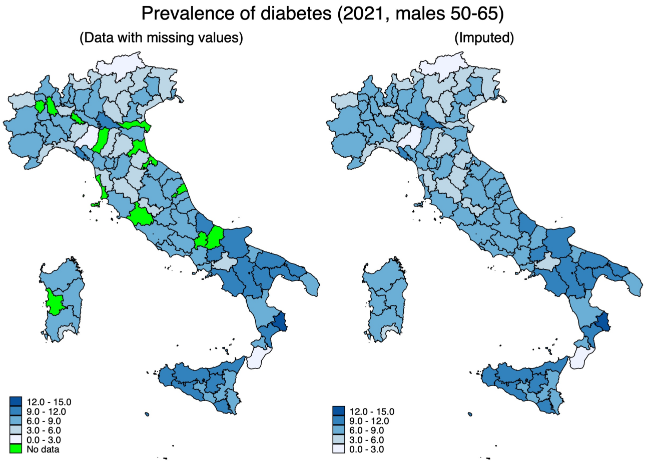
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR Ministero dell'Università e della Ricerca
Finanziato dall'Unione europea NextGenerationEU
GRINS

We are working to incorporate the longitudinal dimension by adopting a dynamic specification of the Durbin model:

$$\boldsymbol{y}_t = \alpha_0 \boldsymbol{y}_{t-1} + \lambda_0 \boldsymbol{W} \boldsymbol{y}_t + \boldsymbol{X}_t \beta_0 + \boldsymbol{W} \boldsymbol{X}_t \gamma_0 + \boldsymbol{\epsilon}_t \quad t = 1, \ldots, T.$$

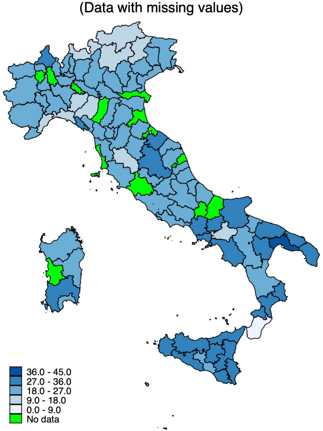
It allows to exploit the serial dependence characterizing most of the chronic conditions.

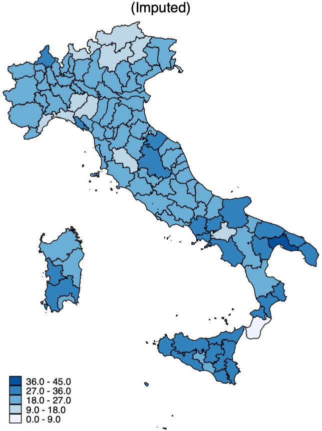
Current approaches (e.g. Zhou et al. 2022, JBES) work only when T is large.

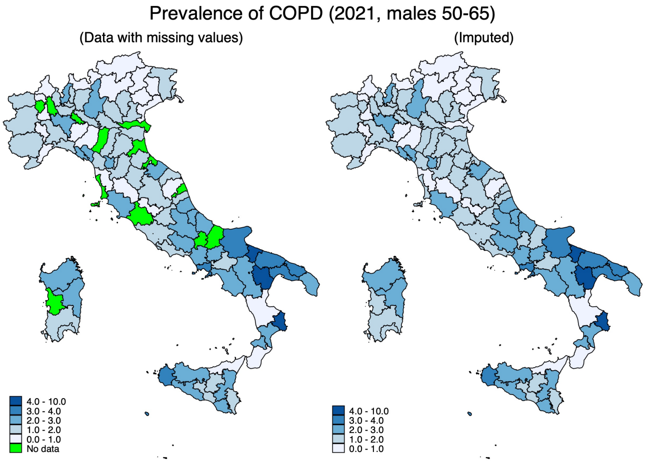

Prevalence of diabetes (2021, males 50-65)
(Data with missing values)
(Imputed)
12.0 - 15.0
9.0 - 12.0
6.0 - 9.0
3.0 - 6.0
0.0 - 3.0
No data
12.0 - 15.0
9.0 - 12.0
6.0 - 9.0
3.0 - 6.0
0.0 - 3.0

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR
Finanziato dall'Unione europea NextGenerationEU
GRINS

Prevalence of hypertension (2021, females 50-65)






Prevalence of COPD (2021, males 50-65)
(Data with missing values)
(Imputed)
4.0 - 10.0
3.0 - 4.0
2.0 - 3.0
1.0 - 2.0
0.0 - 1.0
No data
4.0 - 10.0
3.0 - 4.0
2.0 - 3.0
1.0 - 2.0
0.0 - 1.0

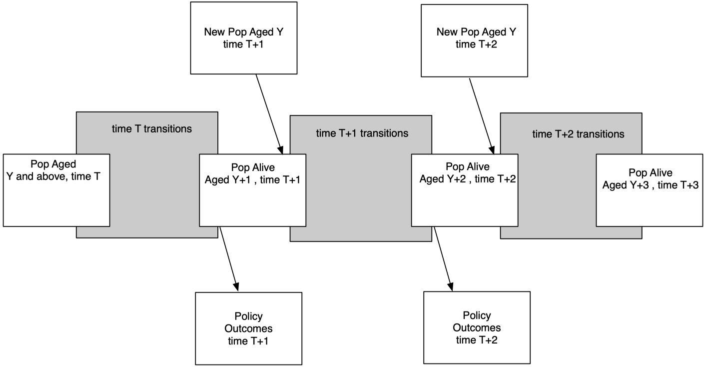
Outline

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR Ministero dell'Università e della Ricerca
Finanziato dall'Unione europea NextGenerationEU
GRINS

- The Future Italian Model is a discrete-time dynamic micro-simulation model able to provide short and medium-run projections (up to 2040).
- It is an adaptation of the well-established Future Elderly Model (FEM) for the Italian population aged 15+.
- Main outcomes:
  1. Prevalence of major chronic conditions (≥ 30) in terms of mortality and expenditure;
  2. Risk factors (BMI, Smoke);
  3. Life expectancy and mortality rates;
  4. Health care utilizations: drugs, diagnostic tests and specialist visits;
  5. Health expenditures.

Model flow

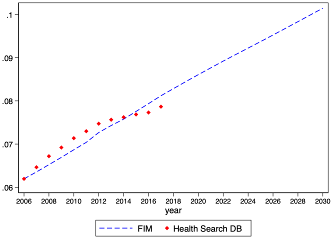
1. Transition module: projects outcomes;
2. Replenishing cohorts module: ensures representativeness of the projections for the population aged 15+;
3. Output module: collects and posts results.

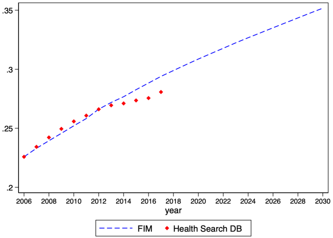


Main data sources

- Health Search Database (HS), a longitudinal observational database run by the Italian College of General Practitioners (SIMG) since 1998.
- ISTAT projections of Italian populations and mortality rate by age, gender and region.
- ISTAT projections of net migration flows by year and region.
- Other variables from different sources that can affect specific diseases (i.e., air pollution when dealing with respiratory diseases).

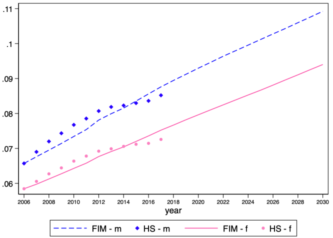
Chronic conditions - I

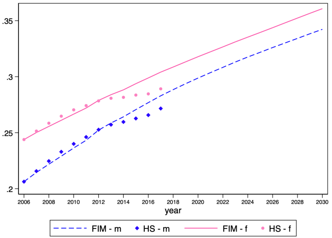




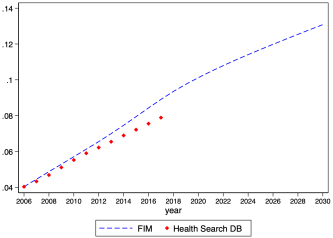
Diabetes

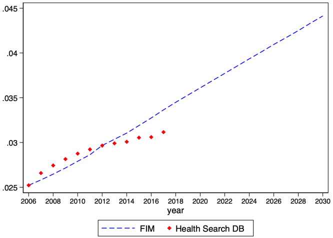
Hypertension





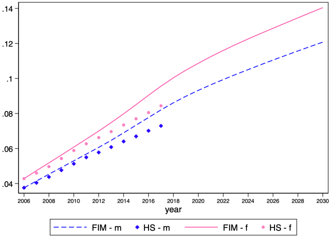
Chronic conditions - II

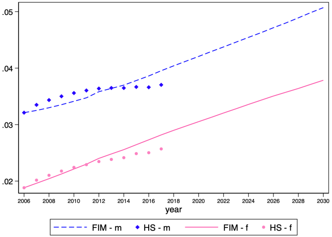




Asthma

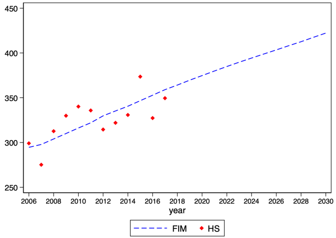
COPD

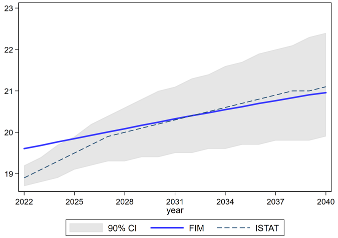




Other outcomes

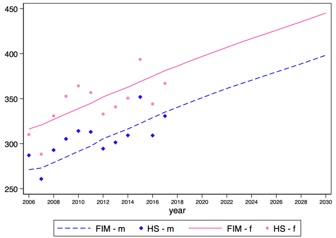
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR
Finanziato dall'Unione europea
NextGenerationEU
GRINS

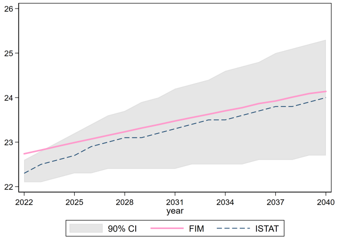




Average total expenditures

Life Expectancy at 65





Outline

Health Search
DASHBOARD

Outline

Why a spin-off?

The healthcare market is changing:

- The industry saw a significant shift towards digital healthcare solutions.
- Healthcare providers and pharmaceutical companies have to navigate extremely complex regulatory landscapes.
- With the increasing focus on preventive healthcare, governments and private healthcare providers are investing in strategies to improve public health outcomes.

There is an increasing demand for the development of cutting-edge, customized analytical and predictive tools to address these issues

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR Ministero dell'Università e della Ricerca
Finanziato dall'Unione europea NextGenerationEU
GRINS

Here's a list sectorial issues that could be addressed with these tools:

- **Demand for Personalized Solutions** The healthcare sector increasingly demanded personalized insights. Generic models, which didn't cater to specific drugs, diseases, or regional healthcare policies, left a significant gap in meeting this demand.
- **Complexity of Healthcare Dynamics**: The healthcare industry, operates in an incredibly complex environment with diverse variables such as drug interactions, patient demographics, and regulatory changes. Off-the-shelf simulation models often struggled to incorporate all these complexities accurately.
- **Rapid Technological Advancements**: The field of microsimulation and data analytics saw continuous advancements. However, the integration of these advancements into practical, customizable tools for specific industry needs often lagged behind the theoretical progress.
- **Growing Recognition of Simulation's Value**: As industries recognized the value of simulation models in decision-making processes, the need for more sophisticated and customizable models became apparent.
- **Varied Policy Landscapes**: Healthcare policies and regulations vary widely across regions. Customizable models that could adapt to these differences were essential for accurate predictions.

When the Spin-off?

- It would be ideal to have a fully operating new enterprise starting before the end of the GRINS project (Feb 2026). Hopefully, it should start operating in January 2025.
- To this aim we need to have completed our work by that time.
- The plan to complete the project is the following:
  - Complete the minimal database to allow the FIM model run by June 2024;
  - Complete the FIM model with all validation and sensitivity analyses by October 2024.

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR
Finanziato dall'Unione europea
NextGenerationEU
GRINS

- **Objective**: Assess the impact of dual therapies with SGLT-2i or DPP-4i type drugs on mortality and cardiovascular diseases.
- We select a cohort of 63,812 diabetic Italians (type 2) in 2017 with no insulin therapy or triple drug therapies.

Baseline cohort characteristics

Variable	Mean	Std. Dev.
Age (years)	71.307	12.159
Females	0.482	0.5
Hypertension	0.768	0.422
Ictus	0.165	0.371
Heart failure	0.069	0.254
Severe hypoglycemia	0.008	0.09
Chronic cardiac ischemia	0.138	0.345
Angina Pectoris	0.028	0.165
Heart attack	0.068	0.252
No therapies	0.017	0.129
Monotherapy	0.46	0.498
Dual therapy	0.523	0.499
Dual therapy with SGLT-2i	0.006	0.076
Dual therapy with DPP-4i	0.027	0.163
Average years from diabetes diagnosis	10.395	6.566

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
MUR
Finanziato dall'Unione europea
NextGenerationEU
GRINS

- **Objective**: Assess the impact of dual therapies with SGLT-2i or DPP-4i type drugs on mortality and cardiovascular diseases.
- We select a cohort of 63,812 diabetic Italians (type 2) in 2017 with no insulin therapy or triple drug therapies.

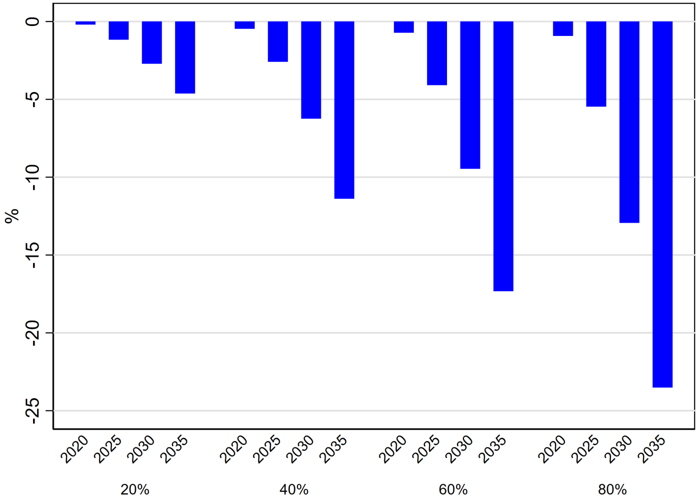
Baseline cohort characteristics

Variable	Mean	Std. Dev.
Age (years)	71.307	12.159
Females	0.482	0.5
Hypertension	0.768	0.422
Ictus	0.165	0.371
Heart failure	0.069	0.254
Severe hypoglycemia	0.008	0.09
Chronic cardiac ischemia	0.138	0.345
Angina Pectoris	0.028	0.165
Heart attack	0.068	0.252
No therapies	0.017	0.129
Monotherapy	0.46	0.498
Dual therapy	0.523	0.499
Dual therapy with SGLT-2i	0.006	0.076
Dual therapy with DPP-4i	0.027	0.163
Average years from diabetes diagnosis	10.395	6.566

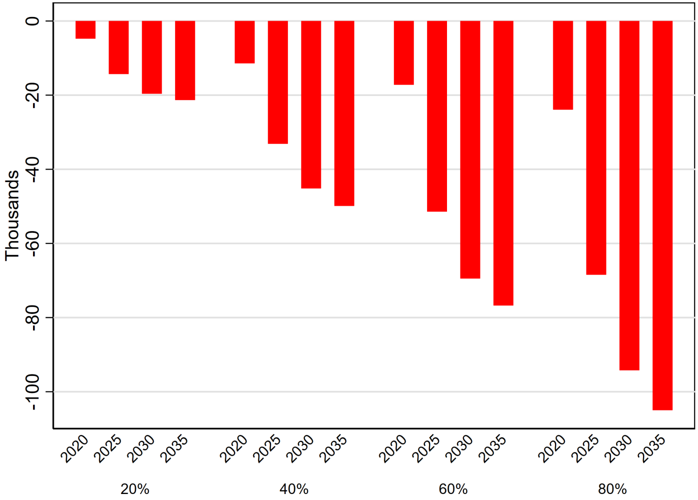
Mortality and CVD

- Reduction in CVD risk (O'Brien et al., 2018):
  1. Dual therapy with DPP-4i: - 27%;
  2. Dual therapy with SGLT-2i: - 38%.
- Reduction in mortality risk:
  1. Dual therapy with DPP-4i: - 35% (Eriksson et al., 2016);
  2. Dual therapy with SGLT-2i: - 32% (Zinman et al., 2015).
- Simulated scenarios (in all scenarios, 70% of those receiving dual therapies receive a DPP-4i while the remaining 30% a SGLT-2i):
  1. 20% of dual therapy based on DPP-4i or SGLT-2i;
  2. 40% of dual therapy based on DPP-4i or SGLT-2i;
  3. 60% of dual therapy based on DPP-4i or SGLT-2i;
  4. 80% of dual therapy based on DPP-4i or SGLT-2i.

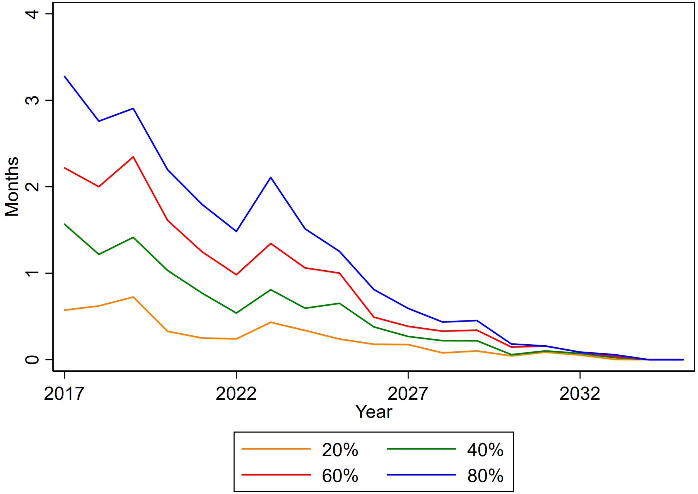
Cumulative reduction in the prevalence of cardiovascular diseases compared to the Status quo scenario



Cumulative premature deaths avoided compared to the Status quo scenario



Gain in life expectancy at 65 compared to the Status quo scenario



Thank you!

GRINS
FOUNDATION