

Towards Fairer Retirement Outcomes: Health-Related Mortality Modelling

Presenters: Andrés Villegas (UNSW)

Co-authors: Francis Hui (ANU), Fei Huang (UNSW), Pramo Samarasinghe (UNSW)

We thank the Australian Government Actuary for their invaluable support in completing this project.



Contents

- 01 Motivation and Background
- 02 Medical Variables
- 03 Model
- 04 How to use the model
- 05 Results
- 06 Future Work and Resources

Motivation and Background

01

Motivation



- The Australian superannuation system's retirement phase needs improvement to better supplement or replace the Age Pension.



- Few Australians invest in longevity products, leading to limited mortality data.



- The Australian Life Table (ALT) lacks detail beyond age and gender.



- This builds on previous work incorporating socio-economic variables.



Goal: Investigate mortality differences by medical and socio-economic variables to improve longevity pool design and retirement outcomes.

Socio-economic Datasets

- Data sourced from Person Level Integrated Data Asset (PLIDA) via ABS DataLab.
- Collaborators: ABS, ATO, Department of Education, Health, Human Services, and Social Services.

Datasets for this Research

- **2016 Census Data:** Entire population snapshot as of August 9, 2016.
- **2009–2016 Social Security Related Information (SSRI):** Benefits, payments, marital status, education.
- **Demographics:** For individuals aged 50+ (post-Dec 2007). Includes gender, birth/death dates, indigenous status.

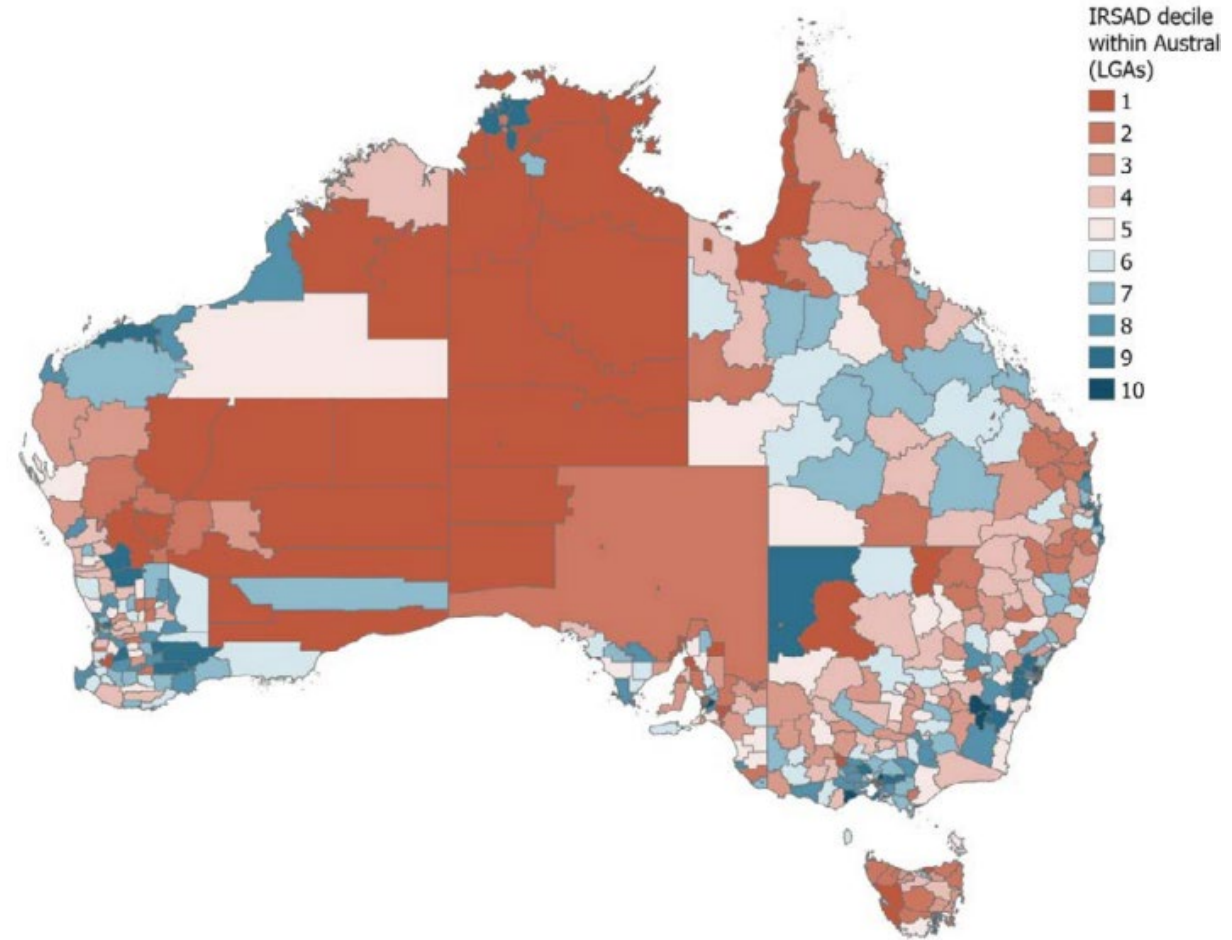
Medical Datasets

- In addition to the previous list;
 - **Medicare Benefits Schedule (MBS):** Healthcare services and utilisation
 - **Pharmaceutical Benefits Scheme (PBS):** Prescription medications and dispensing records
- 2011-2016
- MBS PBS variables: **num** (number of services), **script** (number of scripts dispensed), **qty** (quantity of units supplied), **date**: diagnosis date
- **No diagnostics data**
- Numerous fields with no entries (including chemo, radiation oncology procedures)
- Zero var columns: omitted 76 variables

Socio-economic variables

- Age
- Sex
- Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)
 - one of the SEIFA variables
- Marital Status
- Income
- Homeownership

Note: Education and Occupation are not considered as they are noisy and contains significant missing data for the retirement age population. SEIFA already contains information about education and employment.



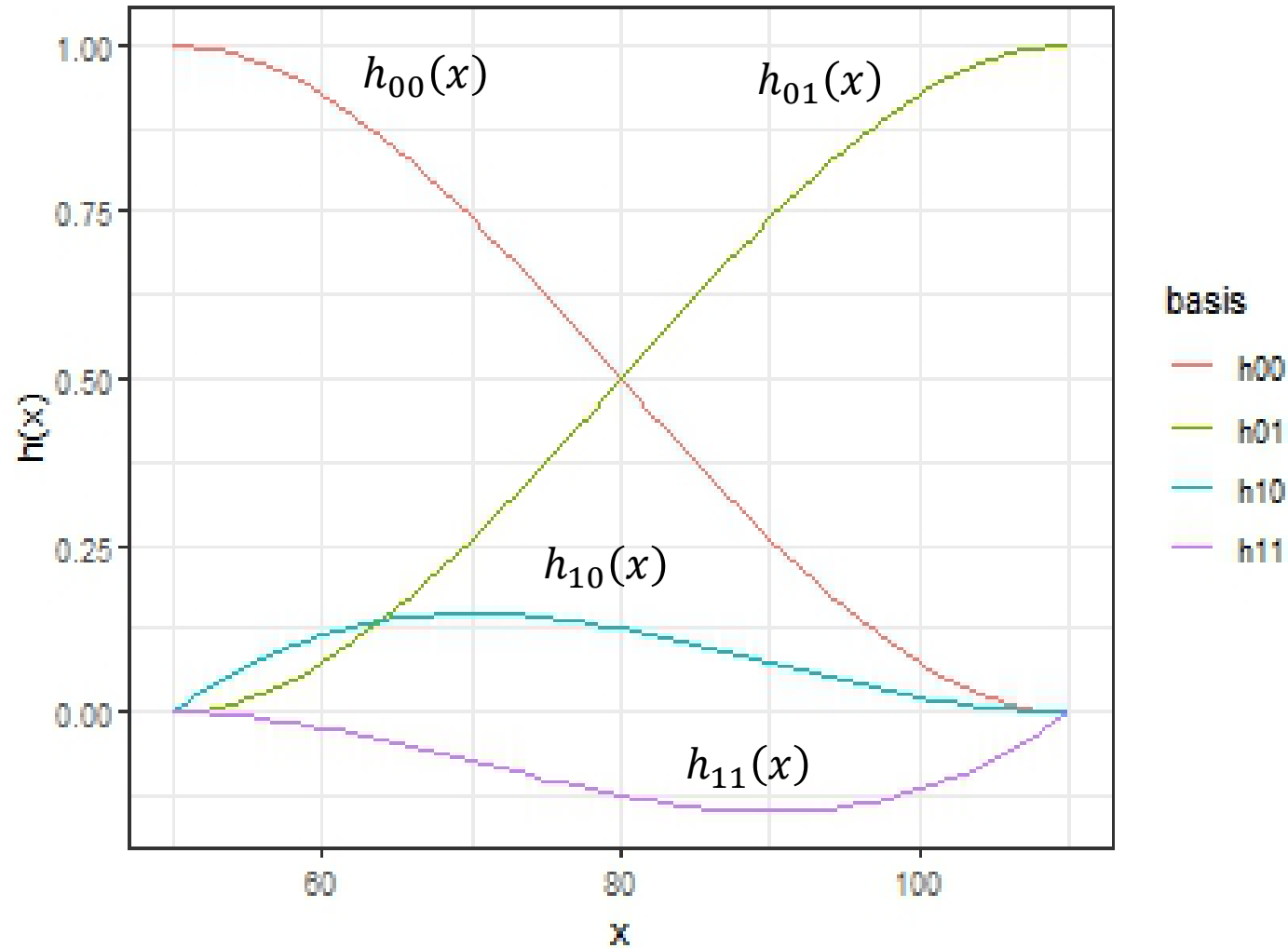
Source: BCARR analysis of the ABS Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD), 2021.

Socio-economic variables

Demographic Variable	Categories	Actuarial and Clinical Proxy
IRSAD	Deciles 1–10	Community wealth, neighbourhood stability, and local healthcare access.
Weekly Personal Income	<\$499, \$500–\$999, \$1000+	Liquid financial capacity to fund care, specialists, and medical payments.
Home ownership Status	Yes, No	Illiquid wealth, baseline financial stability, and long-term housing security.
Marital Status	Married, Single	Social support networks and informal at-home caregiving capacity.
Age	$x \in \{60..110\}$	
Gender	Female, Male	

Hermite splines Model (Richards, 2020)

$$\log \mu_x = ah_{00}(x) + \omega h_{01}(x) + m_0 h_{10}(x) + m_1 h_{11}(x)$$



Note: This is a simplified representation of the model formula. For detailed formulation, please refer to the paper.

Proposed Model for Socio-economic Variables.

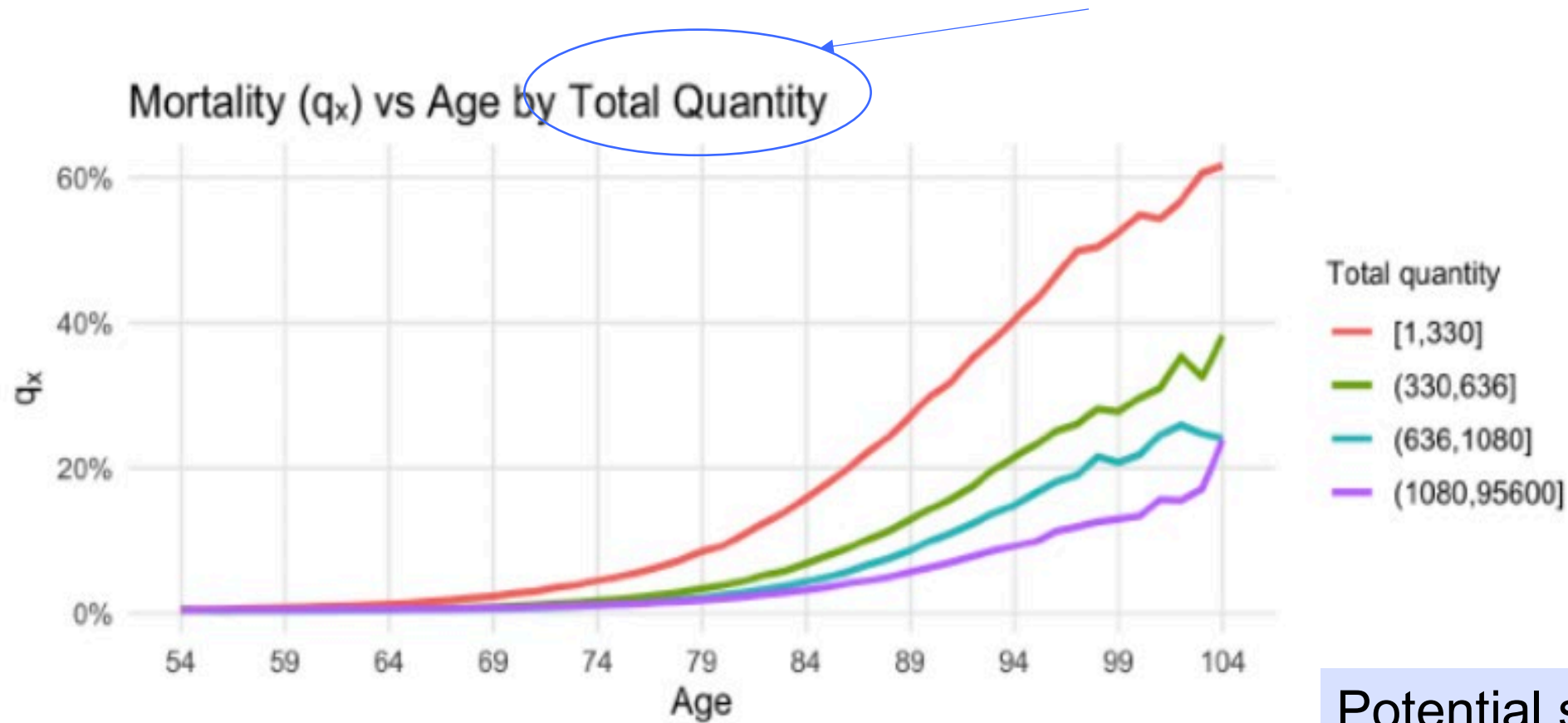
$$\log \mu_{x,s,m,i,h} = ah_{00}(x) + \omega h_{01}(x) + m_0 h_{10}(x) + m_1 h_{11}(x) +$$

$$\left(\begin{array}{c} IRSAD (s) \\ + \\ Marital Status (m) \\ + \\ Income (i) \\ + \\ Homewonership (h) \end{array} \right) \times h_{00}(x)$$

Medical Variables

Information Leakage in MBS/PBS Variables

Total number of units supplied



Potential solutions:

1. Use lagged variables
2. Converting to indicator variables

Medical variables: Summarised

Medical Variable	Actual Codes Used (ATC / MBS)
Diabetes	PBS ATC: A10 (Drugs used in Diabetes) diab_mellitus_num
Parkinson's Disease	PBS ATC: N04 (Anti-Parkinson drugs)
Mental Health	MBS: Mental_Health_num
Antithrombotic Use	PBS ATC: B01AE, B01AF, or B01A (scripts/qty). (Antithrombotic agents)
Pain Medication	MBS: pain_med_num
Cardiac Conditions	PBS ATC: any of C07, C08, C09, C02A, C04, C05A, C01A- (_scripts or _qty).
Coronary Bypass	MBS Items: cor_bypass_num
Lipid Disorders	PBS ATC: C10 (Lipid modifying agents)



Benefits

1. Interpretable
2. Individuals will be able to provide information
3. Less privacy issues due to broader questions.



Disadvantages

1. May result in incorrect classification
(beta blockers used for cardiac and migraines¹)
2. Over-simplification: no intensity captured
3. Due to data limitations, more severe diseases aren't included

¹<https://www.migraine.org.au/beta-blockers>

Medical variables: Summarised

Medical Variable	Actual Codes Used (ATC / MBS)
Diabetes	PBS ATC: A10 (Drugs used in Diabetes) diab_mellitus_num
Parkinson's Disease	PBS ATC: N04 (Anti-Parkinson drugs)
Mental Health	MBS: Mental_Health_num
Antithrombotic Use	PBS ATC: B01AE, B01AF, or B01A (scripts/qty). (Antithrombotic agents)
Pain Medication	MBS: pain_med_num
Cardiac Conditions	PBS ATC: any of C07, C08, C09, C02A, C04, C05A, C01A- (_scripts or _qty).
Coronary Bypass	MBS Items: cor_bypass_num
Lipid Disorders	PBS ATC: C10 (Lipid modifying agents)

Benefits

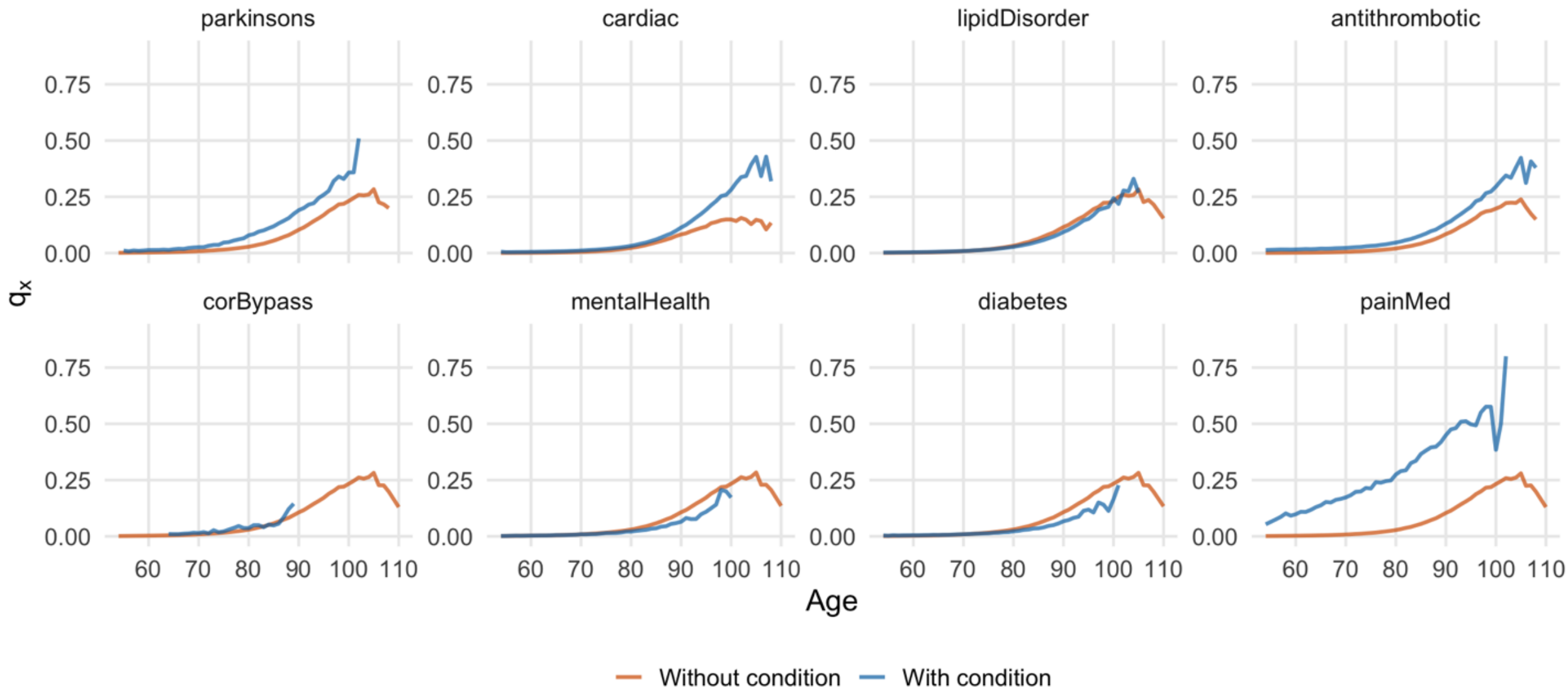
1. Interpretable
2. Individuals will be able to provide information
3. Less privacy issues due to broader questions.

Disadvantages

May result in incorrect classification (beta blockers used for cardiac and migraines¹)
Over-simplification: no intensity captured
Due to data limitations more severe diseases aren't included

¹https://www.migraine.org.au/beta_blockers

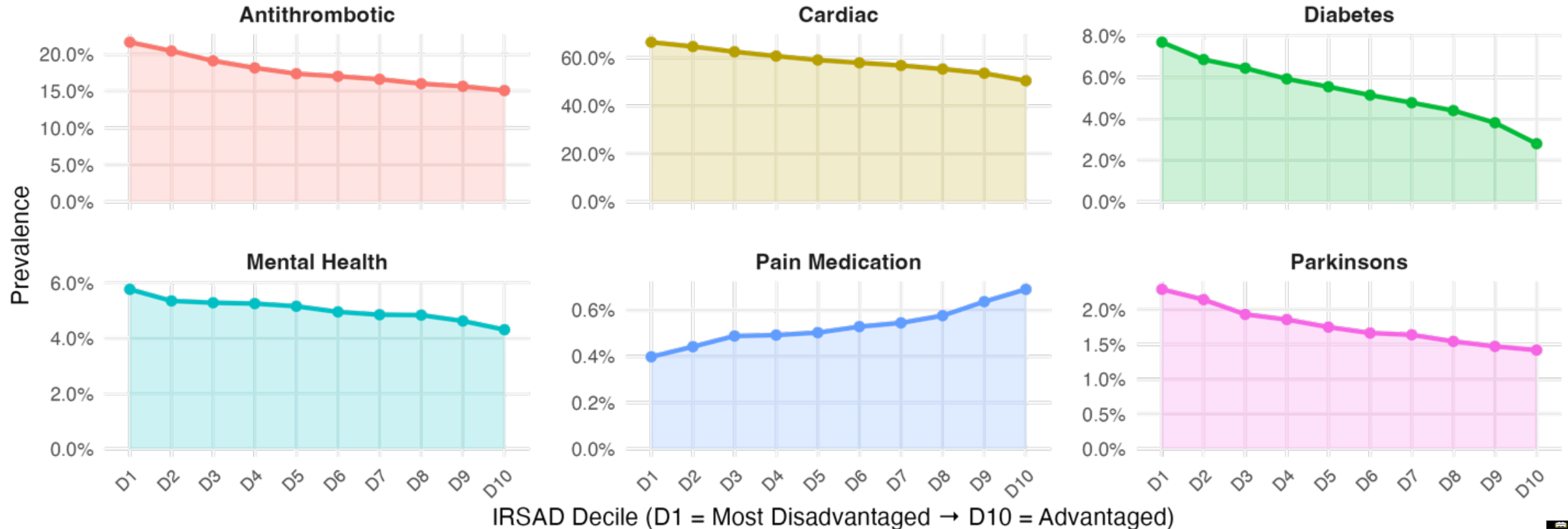
Mortality Rate (q_x) by Age and Condition



Interaction between Medicare and Socio-economic Variables

Health Conditions by Socio-economic Disadvantage (IRSAD)

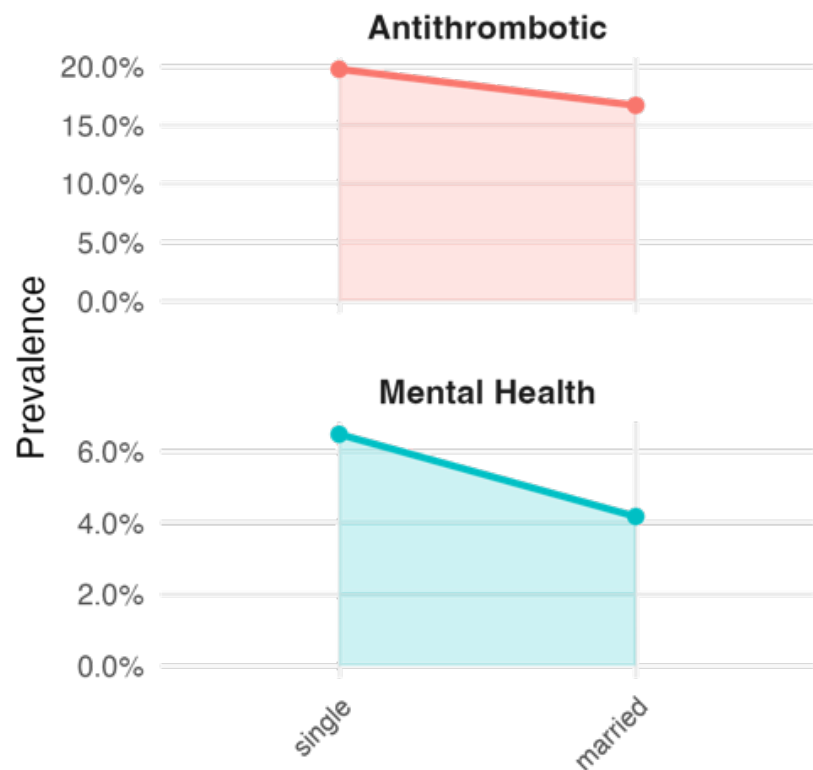
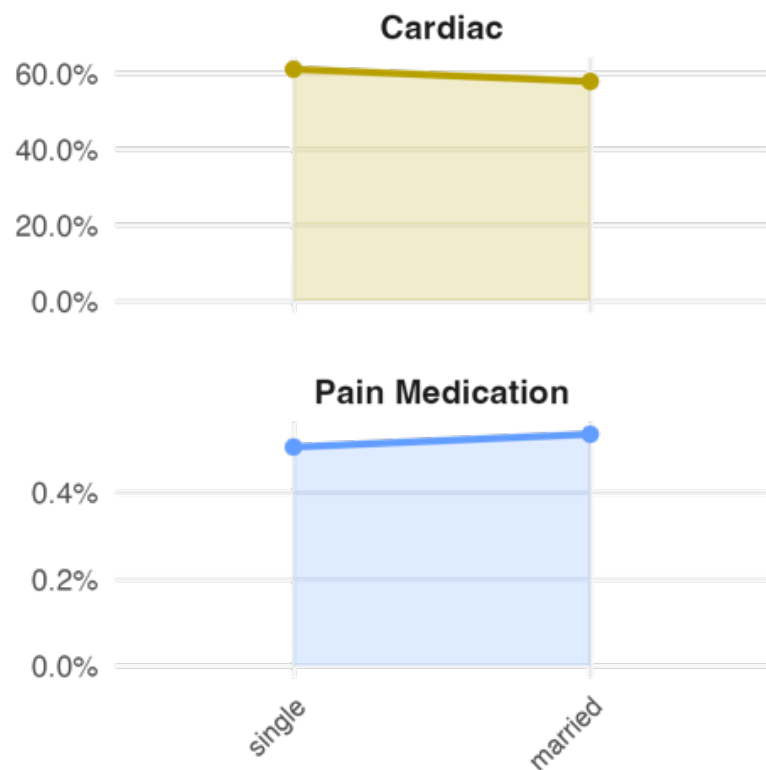
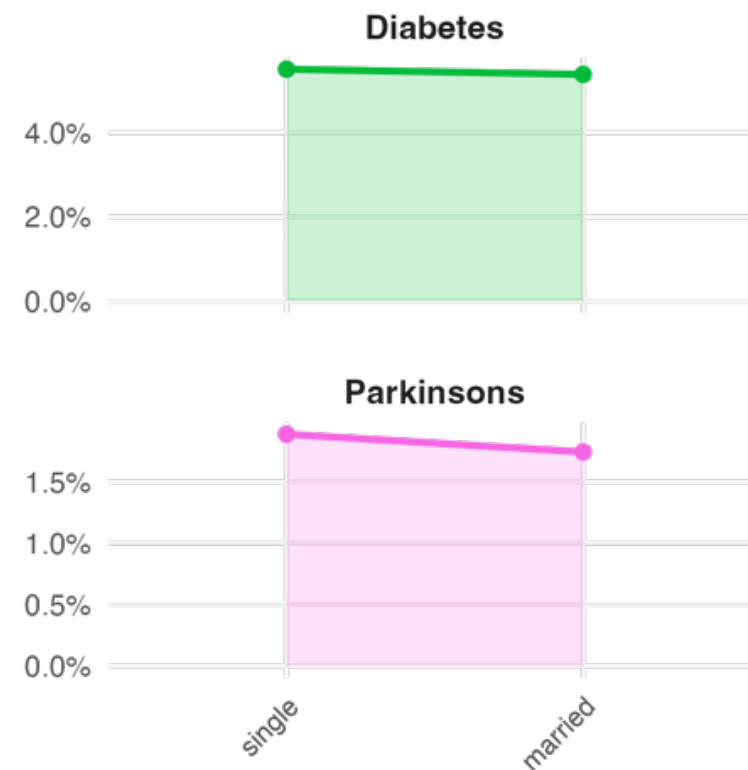
Y-axes scale independently to highlight trend shapes.



Interaction between Medicare and Socio-economic variables

Health Conditions by Marital Status

Y-axes scale independently to highlight trend shapes.

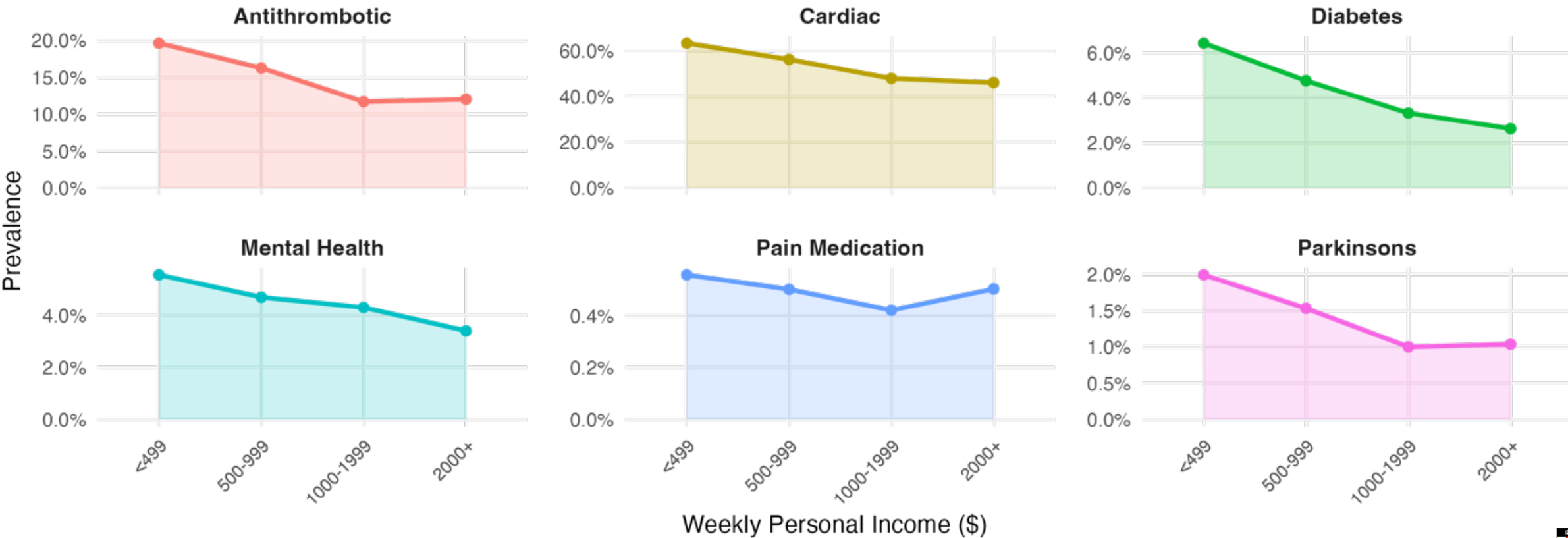


Marital Status

Interaction between Medicare and Socio-economic variables

Health Conditions by Weekly Personal Income Group

Y-axes scale independently to highlight trend shapes.



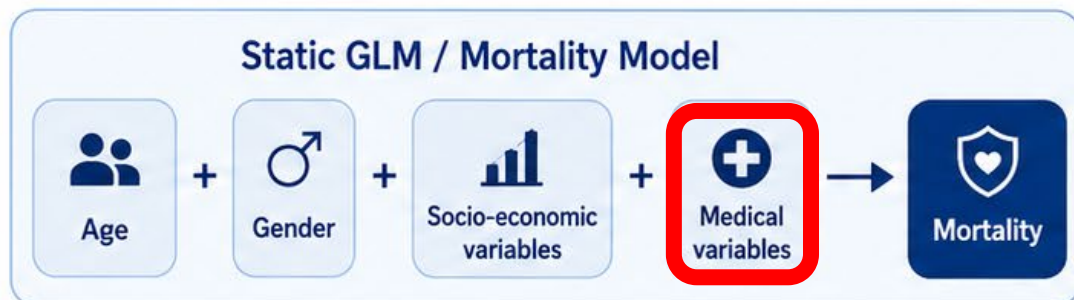


Model

03

Why is another model needed?

Can we just add medical variables to the ALT / demographic model?



Medical conditions are highly correlated

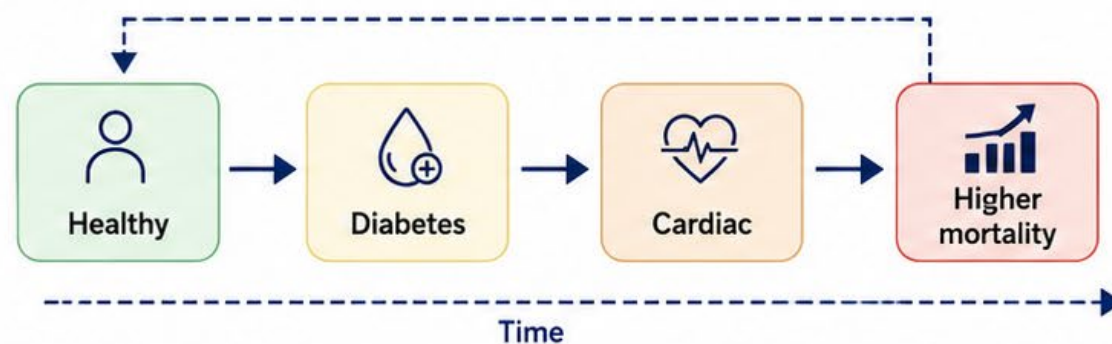


Comorbidity is hard to represent clearly



This treats health as static

What is missing?



Health can worsen or improve over time



Future mortality depends on future health states



A static model cannot capture transitions

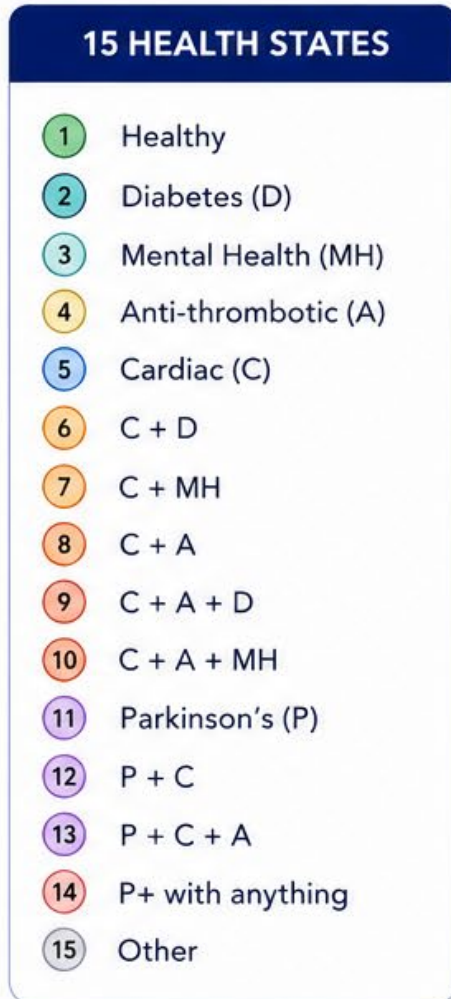
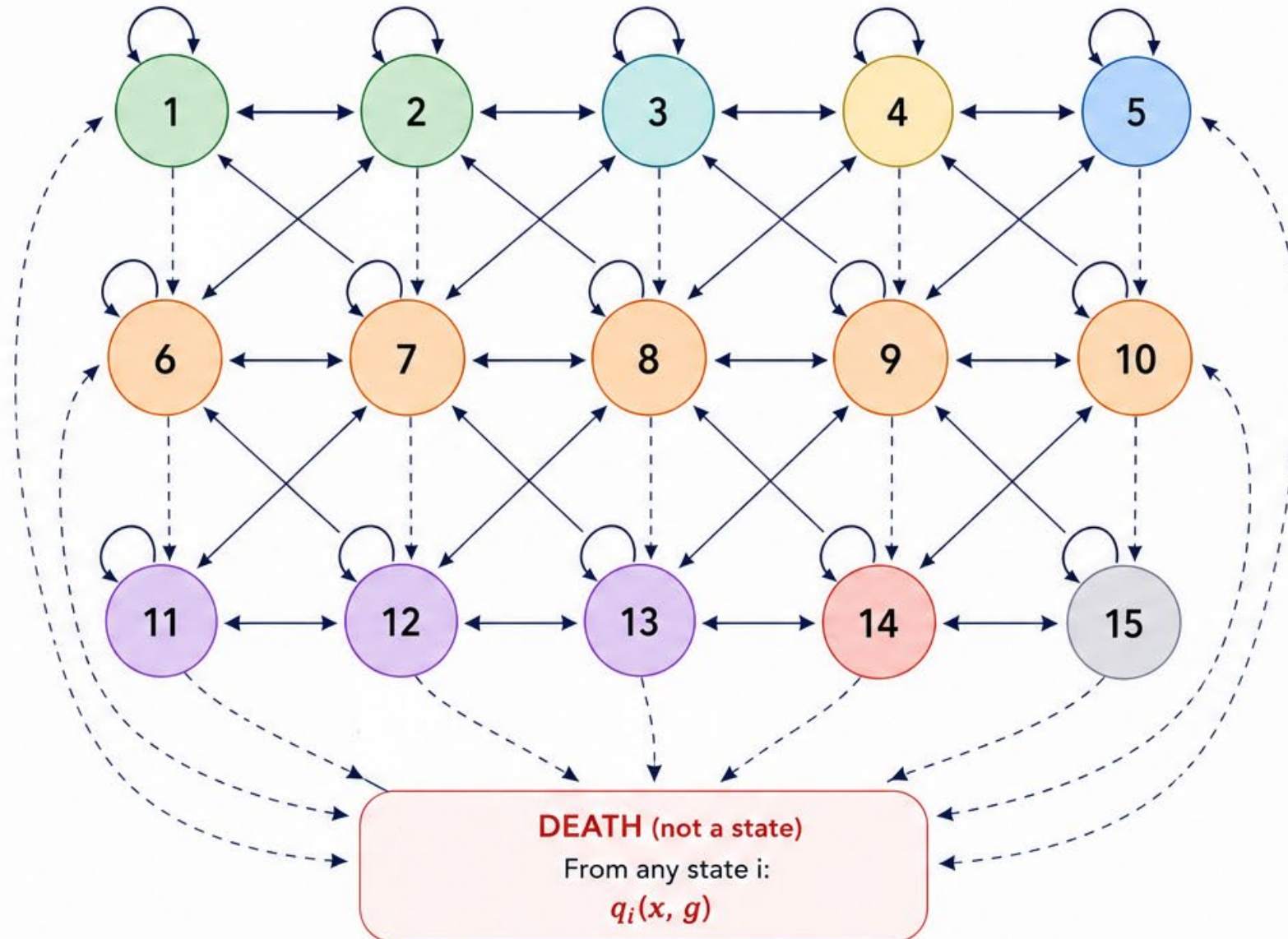


Therefore

We need a dynamic state-based model to capture changing health and its effect on mortality.

Health State–Based Model Using Medical Variable Clusters

Individuals transition between health states; mortality is evaluated within each state (not a state).

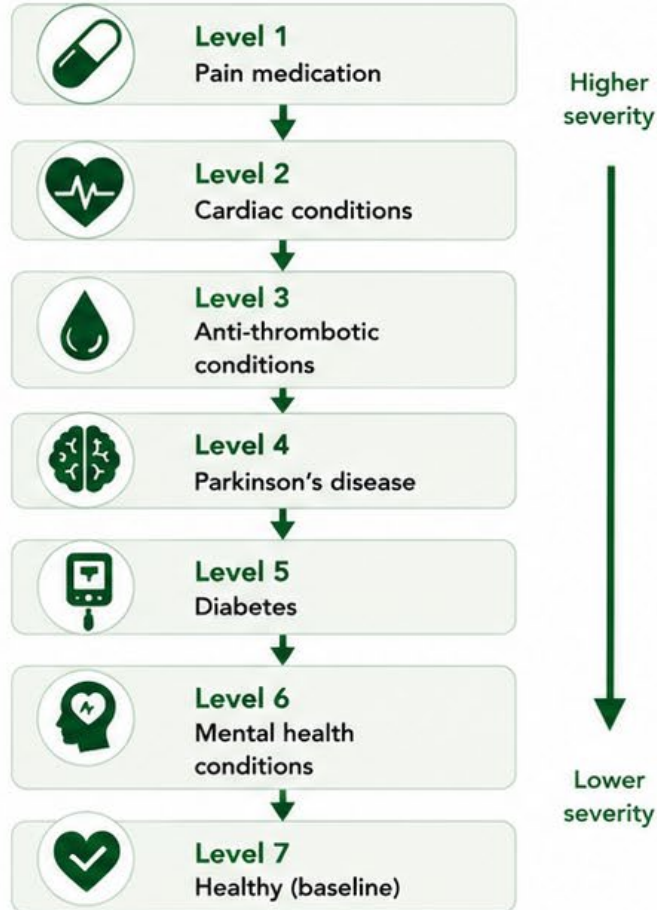


Integrated Medical & Socio-Economic Model

Hierarchy determines the health state cluster; individuals transition between states.

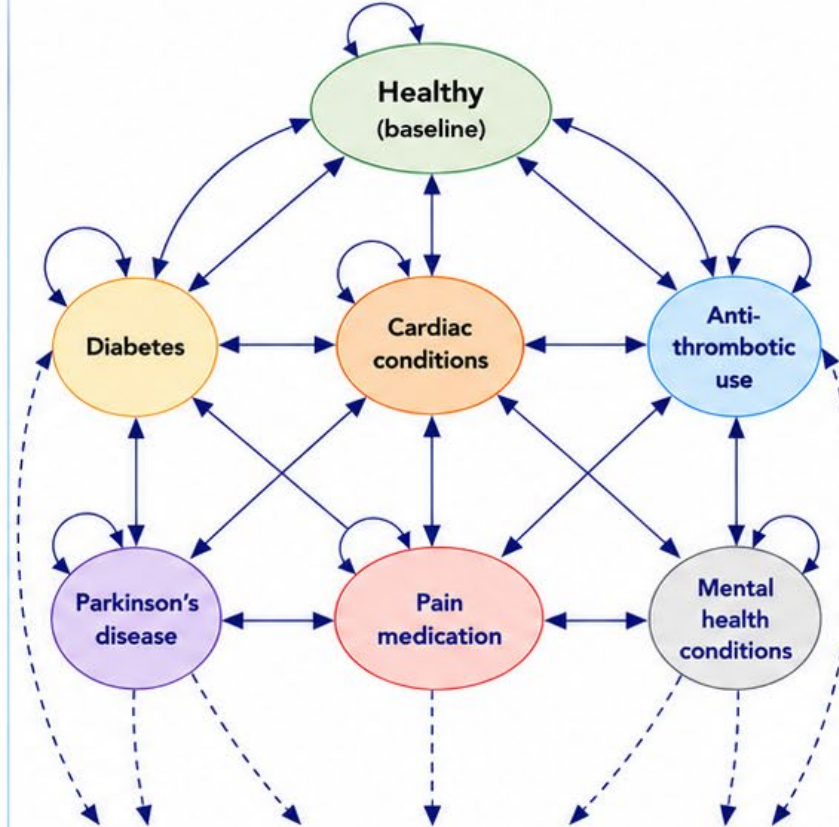
Mortality is evaluated within each state using a state-specific socio-economic model.

1 Hierarchical Classification System (to identify health state cluster)



Assign to the highest applicable level
→ define health state cluster

2 State-Based Markov Model (Health State Clusters)



Mortality evaluated within each state (not a state)

From any state i , probability of death in the next period:

$$q_i(x, g, z) = 1 - \exp\{-\mu_i(x, g, z)\}$$

3 State-Specific Mortality Model (within each health state)

In state i , force of mortality $\mu_i(x, g, z)$ is modelled using cubic plane function:

$$\begin{aligned} \log \mu_i(x, g, z) = & \alpha_{i0} h_{00}(t) + m_{i0} h_{10}(t) + \omega_i h_{01}(t) + m_{i1} h_{11}(t) \\ & + \left(\sum_d \beta_{id} z_d^{\text{IRSAD}} + \sum_k \beta_{ik} z_k^{\text{Income}} \right) \\ & + \left(\sum_h \beta_{ih} z_h^{\text{HomeOwn}} + \beta_{im} z^{\text{Married}} \right) h_{00}(t) \end{aligned}$$

where $h_{00}, h_{10}, h_{01}, h_{11}$ are cubic Hermite basis functions of scaled age $t \in [0, 1]$.

Socio-economic variables (z)

- IRSAD decile (1–10)
- Income level
- Home ownership
- Marital status



Coefficients $\alpha_{i0}, m_{i0}, \omega_i, m_{i1}, \beta_i$ are estimated from the demographic dataset for the subset of individuals in state i .

Model Comparison

Incorporating Medical Variables in Mortality Modelling

Why our proposed approach is better

Adding Medical Variables as a Covariate (in the socio-economic model)

BENEFITS

-  Simple to implement
-  Directly estimates effect of medical variables
-  Uses entire population in one model

DRAWBACKS

-  Assumes same mortality effect across all health statuses
-  Harder to capture disease progression and interactions
-  Lower accuracy and interpretability

Proposed Model (medical variables define states; mortality within each state)

BENEFITS

-  Captures medical complexity via mutually exclusive states
-  Reflects dynamic health progression through transitions
-  More accurate mortality estimates within homogeneous states
-  More interpretable and actionable results

DRAWBACKS

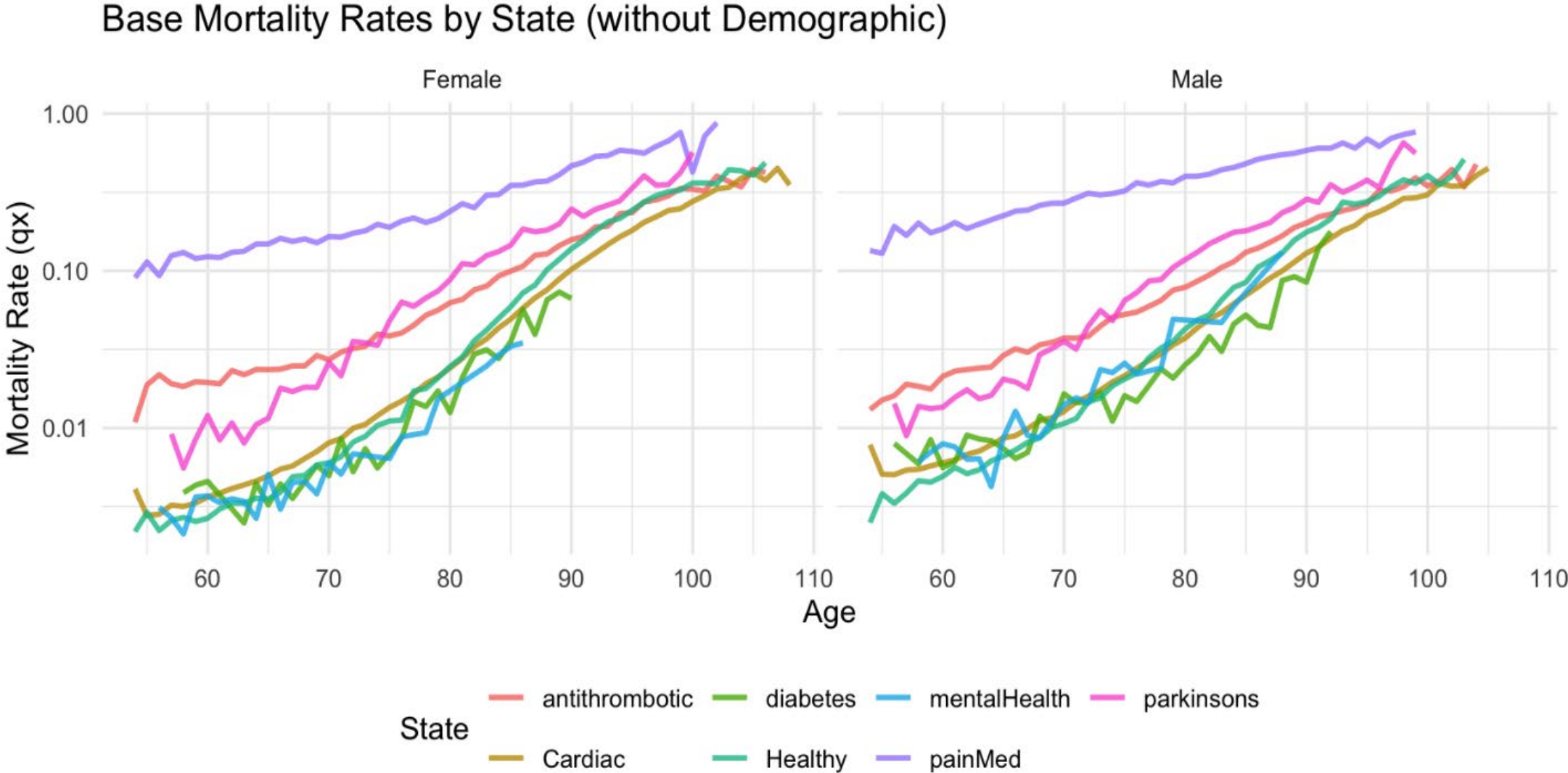
-  More complex to implement
-  Requires sufficient data in each state



Bottom line: Our proposed model better reflects real-world health dynamics, leading to more accurate, interpretable and useful mortality estimates.

How To Use the Model

Mortality per State: Hierarchical System



Coefficients Per Cluster

Finalised Weighted GLM Coefficients - Male

Weighted proportional to observations (n/25,000)

Health State	alpha	m0	omega	m1	IRSAD 1	IRSAD 2	IRSAD 3	IRSAD 4	IRSAD 5	IRSAD 6	IRSAD 7	IRSAD 8	IRSAD 9	Home-owner	Married	Inc \$500-999	Inc \$1000+
healthy	-1.78	-16.22	-0.82	-3.77	-0.03	-0.18	-0.45	-0.54	-0.58	-0.66	-0.69	-0.75	-0.90	-2.29	-0.78	-0.69	-1.19
painmed	-1.91	-8.85	0.01	0.03	0.12	0.02	-0.22	-0.13	-0.12	-0.22	-0.18	-0.14	-0.22	0.19	-0.32	-0.38	-0.72
diabetes	-3.97	-5.54	0.00	4.34	0.09	0.00	-0.15	-0.22	-0.24	-0.31	-0.62	-0.27	-0.87	-1.54	-0.44	-0.41	-0.71
mentalhealth	-4.40	-6.69	-0.06	3.59	0.86	0.74	0.43	0.59	0.52	0.25	0.37	0.22	0.11	2.56	-0.33	-0.60	-1.10
antithrombotic	-0.96	-11.85	-0.37	-0.51	-0.33	-0.59	-0.66	-0.69	-0.71	-0.78	-0.71	-0.92	-1.06	3.47	-0.39	-0.40	-0.70
cardiac	-1.22	-15.82	-0.73	-2.96	-0.32	-0.46	-0.53	-0.67	-0.64	-0.83	-0.83	-0.94	-0.99	-1.88	-0.59	-0.53	-0.93
parkinsons	-5.84	5.00	0.68	8.18	0.52	0.41	0.18	0.19	0.10	0.13	-0.21	0.02	-0.17	-7.37	-0.31	-0.58	-0.94

Finalised Weighted GLM Coefficients - Female

Weighted proportional to observations (n/25,000)

Health State	alpha	m0	omega	m1	IRSAD 1	IRSAD 2	IRSAD 3	IRSAD 4	IRSAD 5	IRSAD 6	IRSAD 7	IRSAD 8	IRSAD 9	Home-owner	Married	Inc \$500-999	Inc \$1000+
healthy	-0.51	-27.98	-1.81	-11.42	-0.28	-0.54	-0.73	-0.79	-0.77	-0.96	-1.02	-1.14	-1.28	-7.20	-0.56	-0.82	-0.98
painmed	-0.75	-14.38	0.10	0.18	-0.08	-0.09	-0.32	-0.12	-0.16	-0.38	-0.15	-0.03	-0.09	-1.74	-0.30	-0.20	-0.41
diabetes	-2.94	-12.52	-0.48	0.74	0.16	0.04	-0.22	-0.30	-0.50	-0.39	-0.30	-0.62	-0.63	-0.01	-0.57	-0.51	-1.12
mentalhealth	-2.55	-17.47	-1.76	-6.85	0.05	0.19	0.05	0.07	-0.19	-0.06	-0.22	-0.21	-0.30	-0.26	-0.40	-0.51	-0.88
antithrombotic	-0.89	-18.90	-0.52	-0.91	0.68	0.49	0.22	0.21	0.30	0.31	0.11	0.10	0.00	1.05	-0.26	-0.42	-0.87
cardiac	-1.01	-19.79	-0.62	-1.63	-0.43	-0.62	-0.77	-0.82	-0.93	-1.00	-1.08	-1.15	-1.35	-4.84	-0.49	-0.54	-0.83
parkinsons	-3.27	-10.72	-1.06	-3.80	0.39	0.17	-0.10	0.07	-0.05	-0.13	-0.34	-0.27	-0.45	-3.69	-0.31	-0.58	-0.97

Model Output: Annual Health-State Transition Matrix

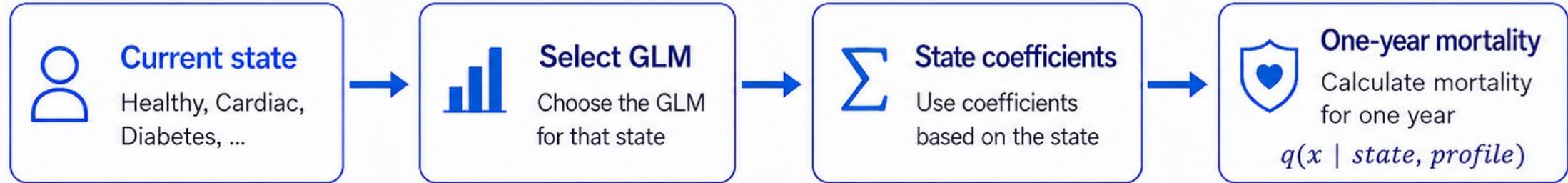


Key Takeaways

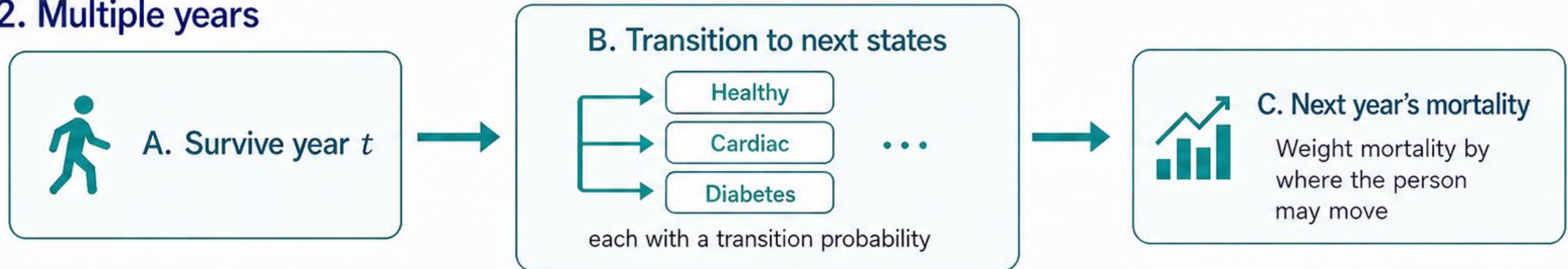
- ✓ **Pain Medication** is highly transient: 25.1% remain, 47.2% move to Cardiac, 20.3% move to Healthy.
- ✓ **Cardiac and Healthy** states are highly persistent (89.6% and 89.8% remain in-state).
With more data, transitions should be modelled in age bands to capture changing risks over time.
- ✓ These transitions **drive the difference** between static and dynamic outcomes.

How the model is used

1. Single year



2. Multiple years



Next-year mortality = survival probability $\times \sum [\text{probability of being in state } j \times \text{mortality in state } j]$



Idea: This replicates a Markov chain calculation, with mortality calculated separately. The only difference is that death is not treated as a state.

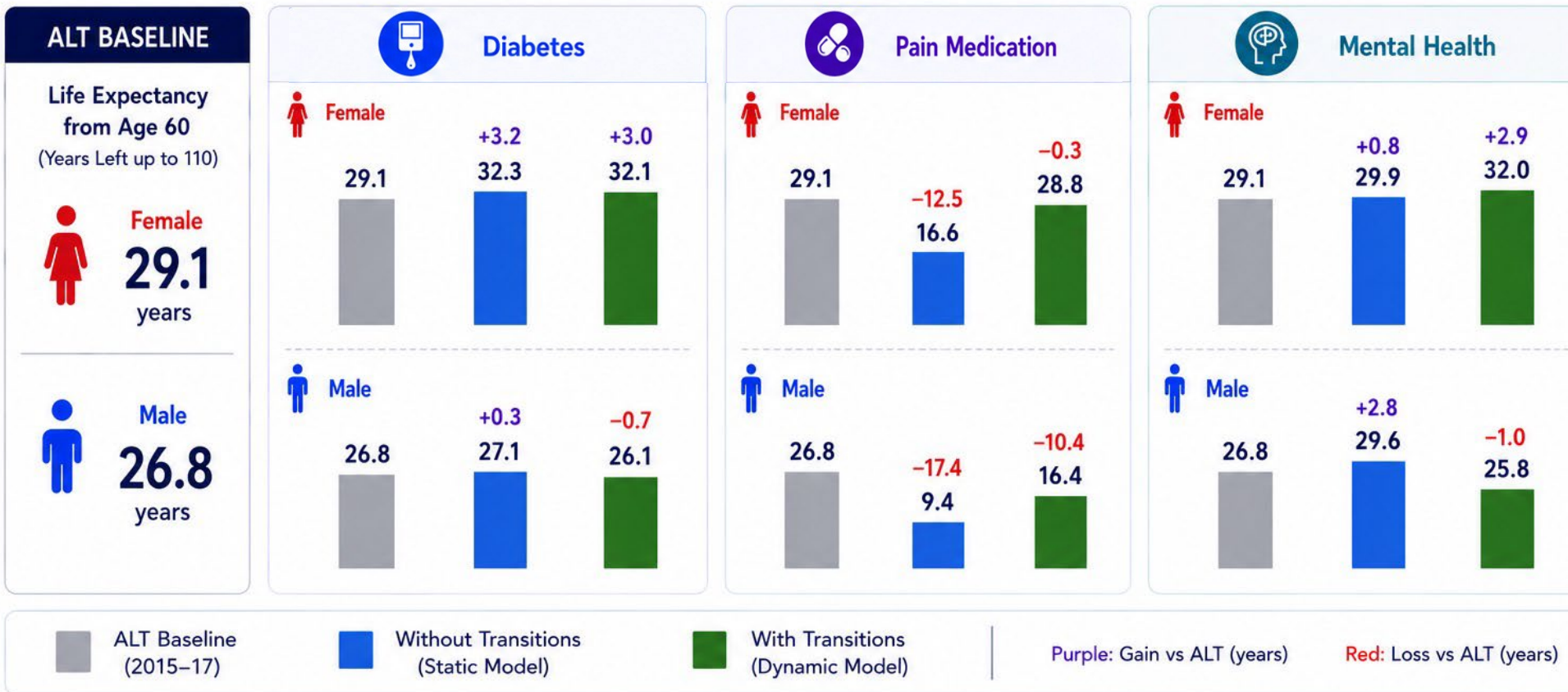
Results

Impact of Health-State Transitions on Life Expectancy (Age 60)

Compared to ALT Baseline (Years Left up to age 110)



Incorporating health-state transitions meaningfully increases life expectancy across all socio-economic profiles. Gains are larger for more disadvantaged groups.



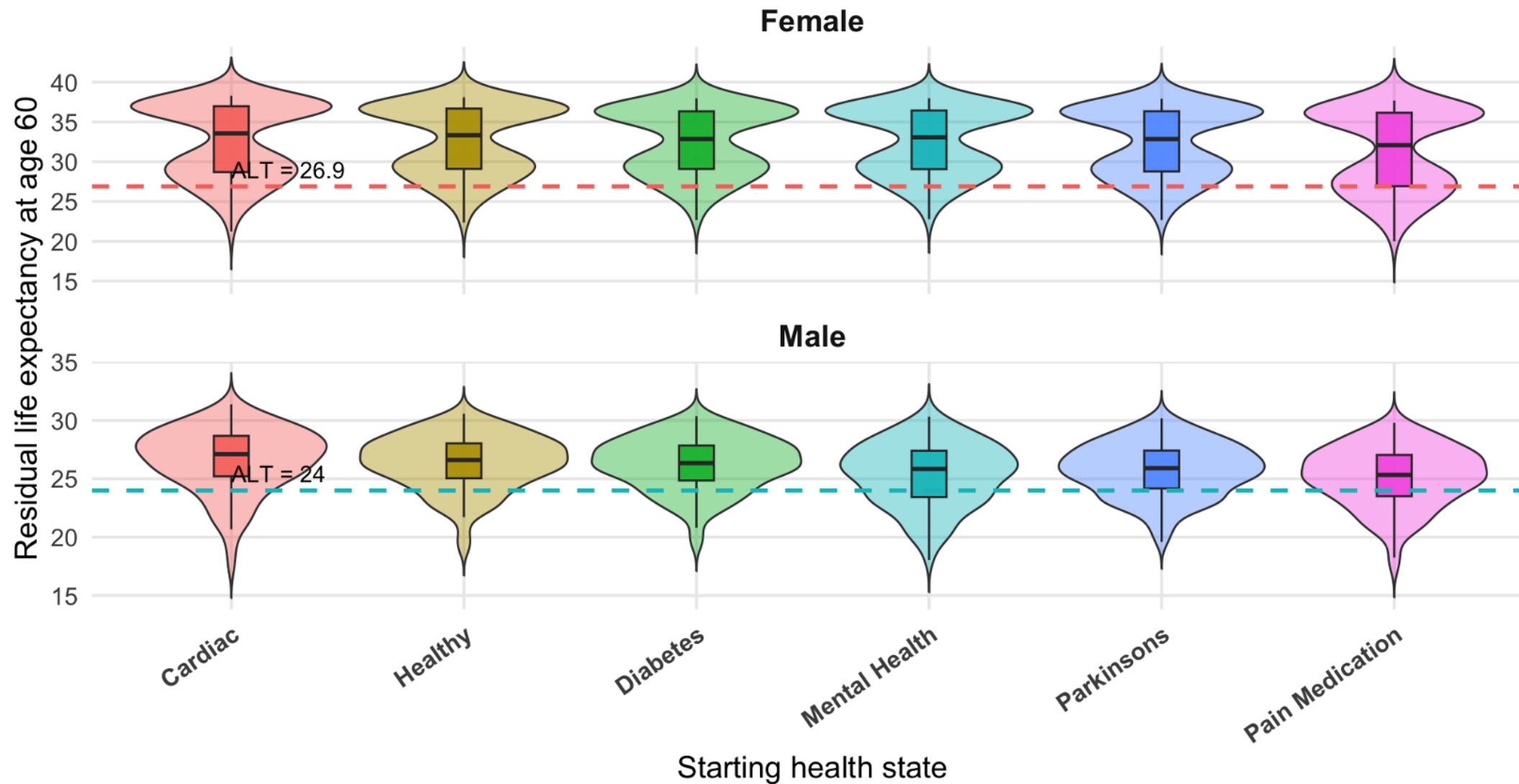
Key Takeaway



Health-state transitions narrow the gap to the ALT baseline, especially for more disadvantaged groups.

Transition-adjusted life expectancy at age 60

Distributions show all socio-economic combinations; dashed line shows ALT



Annuity Income Variation Across Socio-economic Profiles

For a \$100,000 annuity at age 65 (3% interest, no indexation, no death benefit)

Socio-economic model



Female annual income range

\$5,387 – \$6,896



Male annual income range

\$5,785 – \$8,521



A disadvantaged female could receive **28% higher annuity** if socio-economic differences are reflected in pricing.

Improved model



Female annual income range

\$4,694 – \$7,433



Male annual income range

\$5,403 – \$8,510



Under the improved model, a disadvantaged female could receive around **58% higher annuity** if socio-economic differences are reflected in pricing.



ALT 2015–2017 baseline annuity incomes: \$6,321 (female), \$6,973 (male).

Annuity design (e.g., death benefit, indexation, guarantee period) can influence the size of these income differentials.

Future work and Resources

Limitations and Future Improvements



Current Limitations



Short observation window
(2011–2016)



Limited longitudinal tracking
of disease progression



Sparse exposure for rare
disease combinations



Binary indicators only
(no disease severity)



Markov assumption
(depends only on current state)



Missing diagnosis /
cause-of-death data



Limited ability to model
full comorbidity structure



Does not capture comorbidity well
(hierarchical classification)



Future Improvements



Longer longitudinal datasets



Reintroduce severity using
lagged continuous variables



Expand to more granular
disease combinations



Include additional severe conditions
(e.g. cancer, dialysis)



Improve transition modelling
using richer histories



Fully state-specific calibration
(remove weighted population averaging)



More accurate long-term life
expectancy estimation



Relax hierarchical classification
with more data



Key Message

The current model is a **practical first step** toward **dynamic, health-aware** mortality modelling in Australia.



Resources: An Interactive App - “Australian Longevity Explorer”

Australian Longevity Explorer Explorer **Medical Explorer** About Download Data Feedback

RESEARCH PROTOTYPE

Medical Longevity Explorer

Research-use disclaimer: estimates are illustrative, apply only to retired-age profiles above 60, and should not be used as personal medical or financial advice.

How This Component Works

This tab extends the previous socio-economic mortality model by incorporating medical variables and their effect on longevity. The joint model combines demographic, socio-economic, and medical information to estimate mortality, survival, life expectancy, and annuity outcomes.

Reference: [Actuaries Institute research on socio-economic variables and mortality](#)

- Demographic inputs select the profile: age, sex, IRSAD decile, income, and marital status. Home ownership is shown as a provisional field but is currently omitted from the mortality calculation.
- Medical inputs select a starting health state. If multiple conditions are selected, the app uses a single hierarchical state for the mortality calculation.
- The mortality curve uses the embedded coefficient table and the health-state transition matrix to project survival, life expectancy, and annuity payments.

Current limitations: home ownership includes a missing category and remains a developing feature because the relevant data need further refinement. It is currently omitted from the mortality calculation; above age 67, it is forced to missing.

Some medical-state coefficients, especially antithrombotic, can be sensitive to extreme values in this prototype. These estimates should be read

STARTING STATE

Healthy

MODEL LIFE EXPECTANCY (CAPPED AT 110)

19.6 years

ANNUAL ANNUITY INCOME

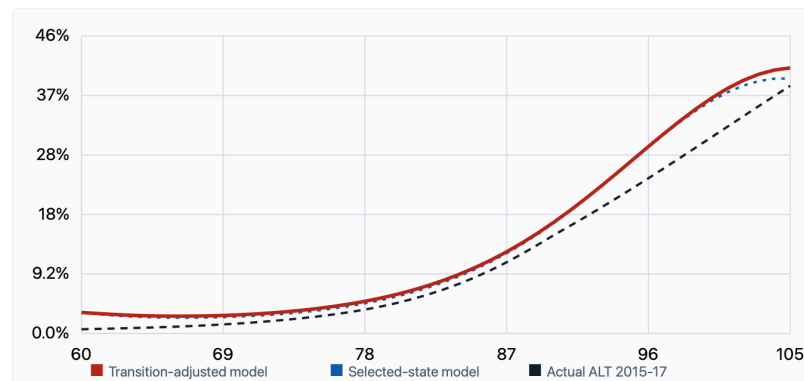
\$7,275

ACTUAL ALT INCOME

\$5,424

Mortality Curve

Annual probability of death by age. The ALT line uses only actual ALT 2015-17 age-sex qx values and does not respond to selected medical conditions.



Socio-economic

Difference in one-year mortality for the current profile compared with the baseline: IRSAD D1, income below \$499, and single. Home ownership is



Thank you!

Questions and discussion

